Dynamical Systems, the Three-Body Problem and Space Mission Design

Wang Sang Koon
California Institute of Technology

Martin W. Lo
Jet Propulsion Laboratory

Jerrold E. Marsden
California Institute of Technology

Shane D. Ross
Virginia Polytechnic Institute and State University

[This version: April 25, 2011]

Mission Design Book Online

This book is available online at
http://www.cds.caltech.edu/~marsden/books/Mission_Design.html,
http://www.cds.caltech.edu/~koon/book/ or
http://www.shaneross.com/books/
with the figures and illustrations in color.

Original Book, v1.0 Completed in 2005 and released in 2006.
Version v1.1 with minor updates was released on, September 25, 2008.
This version v1.2 was revised on April 25, 2011

Wang-Sang Koon: *To my wife Marilyn*

Martin Lo: *To my mother Julia*

Jerry Marsden: *To my niece Kathy*

Shane Ross: *To my wife Jessica*

Contents

Preface

This book was born out of recent exciting activity that is bringing the insights of geometric mechanics and dynamical systems theory to bear on the design of trajectories for space missions already launched or being planned by NASA and ESA.

The pace of space-based achievement is set by capability and affordability. The desire to expand space mission capabilities, and at low cost, has required new approaches to the old problem of space travel. Many recent space missions have complex and highly non-Keplerian orbits that are related to the subtle dynamics of the three- and N-body gravitational problems.

The Future of Space Exploration. For instance, recent initiatives for space exploration involve developing a robust and flexible capability to visit several potential destinations. A Lunar Gateway Station near L_1, a point of balance between the Earth and Moon, has been proposed as a 21st century hub for science and a jumping off point for deep space missions, eventually to land humans on Mars. Furthermore, scientists want to obtain samples from the Moon, asteroids, and Mars, and return them to Earth for analysis to better understand the origin, history, and future of our solar system.

Looking beyond our celestial neighborhood, space agencies envision placing new kinds of telescopes into deep space, such as NASA's Terrestrial Planet Finder and ESA's Darwin project, whose goal would be to find Earth-like planets around other stars, planets which may harbor life.

One need not look beyond our own solar system to find tantalizing hints of life. Less than a billion kilometers away orbiting Jupiter are some of the

most interesting sites of future exploration. The four planet-sized moons of Jupiter—Io, Europa, Ganymede, and Callisto—are fascinating worlds in their own right. These moons, first seen by Galileo Galilei when he invented the telescope four centuries ago, have recently revealed a diversity and complexity that was not expected.

One of these icy moons, Europa, has provided the most evidence that it may have an ocean, perhaps as deep as 100 kilometers. In fact, the ocean may come close enough to the surface at points that a future spacecraft could burrow through the ice layer, deploying a submarine robotic explorer to show us the first pictures of a extraterrestrial ocean. Before such a mission could take place, an exploratory mission to orbit the Galilean moons is necessary for identifying possible landing sites.

The Challenge of Space Mission Design. Space missions which reach these destinations are complex and challenging to design, requiring new and unusual kinds of orbits to meet their goals, orbits that cannot be found by classical approaches to the problem.

Classical approaches to spacecraft trajectory design have been quite successful: for instance, Hohmann transfers for the Apollo Moon landings, and swingbys of the outer planets for Voyager. But these missions were costly in terms of fuel, e.g., large burns for orbit entry were required for Apollo.

The minimization of fuel (i.e., energy) requirements for a spacecraft's trajectory is important for the feasibility of its mission. An unreasonably high fuel requirement can render a mission infeasible.

The Promise of Low Energy Trajectories. Fortunately, a new class of *low energy trajectories* have recently been discovered and employed which make possible missions which classical approaches could not. Such trajectories are achieved by making use of gravity as much as possible, using the natural dynamics arising from the presence of a third body (or more bodies). The term "low-energy" is used to refer to the low fuel and therefore low energy required to control the trajectory from a given starting condition to a targeted final condition.

Low energy trajectory technology allows space agencies to envision missions in the near future involving long duration observations and/or constellations of spacecraft using little fuel. A proper understanding of low energy trajectory technology begins with a study of the restricted three-body problem, a classic problem of astrodynamics.

The Point of View Taken in This Book. This monograph focuses on research conducted by the authors and related topics by co-workers. The general point of view is to consider global solutions to the restricted three-body problem from a geometric point of view. We seek dynamical channels in the *phase space* which wind around the planets and moons and naturally connect them. These *low energy passageways* could slash the amount of fuel spacecraft need to explore and develop our solar system.

In order to effectively exploit these passageways, we will discuss the global transport mechanism in the restricted three-body problem involving the stable and unstable invariant manifolds associated to two *libration points*, equilibrium locations in a rotating two-body gravity field situated along a line joining the two attracting bodies.

Our hope is the book can serve as a valuable resource for graduate students and advanced undergraduates in aerospace engineering, as well as a manual for practitioners who work on libration point and deep space missions in industry and at government laboratories. We include a wealth of background material, but also bring the reader up to a portion of the research frontier. Moreover, students and professionals in dynamical astronomy and applied dynamics will also find material of interest.

We expect that this book will prepare the reader for other in-depth works related to libration point orbit mission design, such as the four volume set *Dynamics and Mission Design Near Libration Points* by Gómez, Jorba, Llibre, Martínez, Masdemont, and Simó [2001].

Furthermore, the book goes beyond the traditional scope of libration point mission design, developing tools for the design of trajectories which take full advantage of natural three or more body dynamics, thereby saving precious fuel and gaining flexibility in mission planning. This is key for the development of some innovative mission trajectories, such as low energy lunar missions and low energy tours of outer planet moon systems, such as a mission to tour and explore in detail the icy moons of Jupiter.

Book Websites. We urge the reader to look at the book's websites: http://www.cds.caltech.edu/~marsden/books/Mission_Design.html, http://www.cds.caltech.edu/~koon/book/ or http://www.shaneross.com/books/space, where convenient and relevant links to literature, software, and updates will be posted.

Acknowledgments. We thank Gerard Gómez and Josep Masdemont for many helpful discussions and for sharing their wonderful software tools with us which made many of the illustrations possible. We thank Kathleen Howell, Brian Barden, and Roby Wilson for the inspirational work on the Genesis Discovery Mission. Much of their work is reflected in this book.

The work which has gone into this book was done over the course of a decade and in that time we have received input from many people. The following is a partial list of colleagues we wish to thank for helpful discussions and comments: Kyle Alfriend, Al Barr, Edward Belbruno, Julia Bell, Erik Bollt, Joel Burdick, Alan Chamberlin, Kwing Lam Chan, Siu Yuen Cheng, Gerald Condon, Fred Culick, Michael Dellnitz, Donald Dichmann, Eusebius Doedel, David Dunham, Robert Farquhar, Mathieu Desbrun, David Farrelly, Dave Folta, Harmon Fowler, Frederic Gabern, Peter Goldreich, Martin Hechler, Bambi Hu, Andrew Ingersoll, Charles Jaffé, Àngel Jorba, Oliver Junge, Jean Kechichian, Herb Keller, Erica Klarreich, Sanjay Lall, Andrew Lange, Francois Lekien, Hal Levison, Robert Little-

john, Jaume Llibre, Brian Marsden, Regina Martínez, Vasile Massimiliano, Harris McClamroch, Richard McGehee, William McLaughlin, James Meiss, Ken Meyer, Rick Moeckel, Alessandro Morbidelli, Richard Murray, Ken Museth, Paul Newton, Michael Ortiz, Kathrin Padberg, Randy Paffenroth, Phil Palmer, Linda Petzold, Robert Preis, Nicole Rappaport, David Richardson, Mark Roberts, Vered Rom-Kedar, Ralph Roncoli, Clarence Rowley, Donald Saari, Daniel Scheeres, Radu Serban, Joel Sercel, Carles Simó, Douglas Smith, David Spencer, Ted Sweetser, Gary Taubes, Bianca Thiere, Scott Tremaine, Turgay Uzer, Luz Vela-Arevalo, Benjamin Villac, Alan Weinstein, Laurent Wiesenfeld, Greg Whiffen, Stephen Wiggins, Daniel Wilczak, Donald Yeomans, Piotr Zgliczyński, Shiqing Zhang, and Jilin Zhou.

The book makes use of research that was originally supported by the Caltech President's Fund in 1999 and 2003, the NASA Advanced Concepts Research Program, The Genesis Project, a Max Planck Research Award, NSF grants KDI/ATM-9873133 and DMS-0402842, NSF/ITR Grant ACI-0204932, and AFOSR Microsat contract F49620-99-1-0190. In addition, parts of the research were carried out at the Jet Propulsion Laboratory and the California Institute of Technology under a contract with the National Aeronautics and Space Administration.

1
Introduction

In the first chapter we present an overview of the main themes of the book, astrodynamics, Hamiltonian dynamical systems and the methods used to compute sets of trajectories in dynamical systems. The chapter is written more in a review style than are the subsequent ones and, hence, serves the main purpose of introducing topics covered more thoroughly later in the book. At the end of the chapter, we outline the topics to be covered in the chapters that follow.

1.1 Astrodynamics and Dynamical Astronomy

Astrodynamics and dynamical astronomy apply the principles of mechanics, including the law of universal gravitation to the determination of the motion of objects in space. Orbits of astronomical bodies, such as planets, asteroids, and comets are calculated, as are spacecraft trajectories, from launch through atmospheric re-entry, including all the needed orbital maneuvers.

While there are no sharp boundaries, astrodynamics has come to denote primarily the design and control of spacecraft trajectories, while dynamical astronomy is concerned with the motion of other bodies in the solar system (origin of the moon, Kuiper belt objects, etc). From a dynamical systems perspective of interest to us, it is quite useful to mix these subjects. There

is one obvious commonality: the model used for studying either a spacecraft or, say, the motion of an asteroid is the restricted $N + 1$ body problem, where N celestial bodies move under the influence of one another and the spacecraft or asteroid moves in the field of these bodies, but has a mass too small to influence their motion.

The Ephemeris and Its Approximations. In the case of motion within the solar system, the motion of the N bodies (planets, moons, etc) can be measured and predicted to great accuracy, producing an ephemeris. An *ephemeris* is simply a listing of positions and velocities of celestial bodies as a function of time with respect to some coordinate system. An ephemeris can be considered as the solution of the N-body gravitational problem, and forms the gravitational field which determines a spacecraft or asteroid's motion.

While the final trajectory design phase of a space mission or the long term trajectory of an asteroid will involve a solution considering the most accurate ephemeris, insight can be achieved by considering simpler, approximate ephemerides (the plural of ephemeris). An example of such an ephemeris is a simplified solution of the N-body problem, where N is small, for example, the motion of the Earth and Moon under their mutual gravitation, a two-body solution. The simplest two-body solution of massive bodies which gives rise to interesting motion for a spacecraft is the circular motion of two bodies around their common center of mass. The problem of the spacecraft's motion is then known as the *circular restricted three-body problem*, or the CR3BP. One goal of this book is to instill the reader with intuition regarding this problem, as well as provide analytical and computational tools for its solution.

Introduction to the Trajectory Design Problem. The set of possible spacecraft trajectories in the three-body problem can be used as building blocks for the design of spacecraft trajectories in the presence of an arbitrary number of bodies. Consider the situation shown in Figure 1.1.1, where we have a spacecraft, approximated as a particle, P, in the gravitational field of N massive bodies. We assume P has a small enough mass that it does not influence the motion of the N massive bodies, which move in prescribed orbits under their mutual gravitational attraction. In the solar system, one can think of a moon, M_2, in orbit around a planet, M_1, which is in orbit around the Sun, M_0.

The goal of trajectory design is to find a transfer trajectory, such as the one shown in Figure 1.1.2(a), which takes the spacecraft from a prescribed initial orbit to a prescribed final orbit using controls. The initial orbit may be an orbit around the Earth and the final orbit an orbit around one of the moons of Jupiter, for instance. To effect this transfer, we could use high thrust or low thrust propulsion systems. In the low thrust case, we have a small continuous control which can operate at all times. In the high thrust case, we assume that the control is discretized into several instantaneous

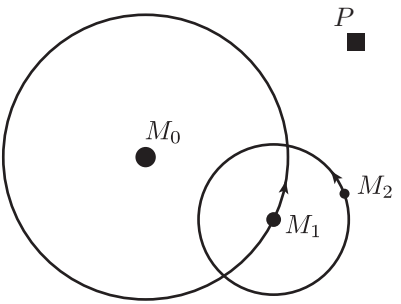


FIGURE 1.1.1. A spacecraft P in the gravitational field of N massive bodies which move in prescribed orbits.

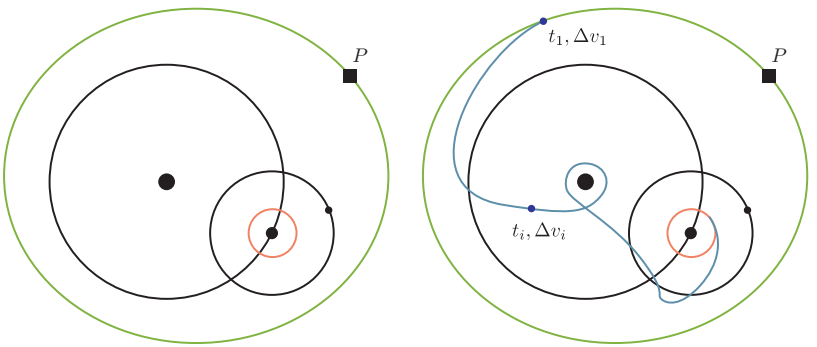


FIGURE 1.1.2. (a) The goal is to find a transfer trajectory which takes the spacecraft from an initial orbit to a final orbit using controls. (b) Assuming impulsive controls, i.e., several instantaneous changes in the spacecraft's velocity, with norm Δv_i at time t_i, we can effect such a transfer.

changes in the spacecraft's velocity. These instantaneous changes have a magnitude Δv_i at time t_i. Under a high thrust assumption, the Δv's are proportional to the fuel consumption:

$$\Delta v = -v_e \frac{\Delta m}{m}$$

where m is the mass of the rocket and Δm is the mass of propellant ejected at an exhaust velocity v_e (Roy [1988]). As spacecraft are limited in the amount of fuel that they can carry on-board for their journey, we often want to consider an *optimal control problem*: minimize the fuel consumed (equivalently, energy). In other words, we want to find the maneuver times t_i and sizes Δv_i to minimize

$$\sum_i \Delta v_i,$$

the total change in velocity, or "ΔV" as it is called.

It is typical in space missions to use the magnitude of the required ΔV as a measure of the spacecraft fuel performance. The propellant mass is a much less stable quantity as a measure of spacecraft performance, since it is dependent on the spacecraft mass and various other parameters which change frequently as the spacecraft is being built. The ΔV comes from astrodynamics considerations only and is independent of the mass and type of spacecraft. Thus, for a given mission objective, one generally wants to *minimize* ΔV.

1.2 The Patched Three-Body Approximation

To get a spacecraft from, say, Earth to other parts of the solar system, it is necessary to find solutions for the motion of the spacecraft under the influence of N bodies, a notoriously difficult problem. Furthermore, one needs to find solutions with a desired behavior, e.g., flying by the giant outer planets as Voyagers 1 and 2 did, while satisfying engineering constraints, e.g., low fuel consumption, short time of flight, low radiation dose, etc.

The Patched Conic Approximation and the Voyager Trajectory. For many purposes it is satisfactory to simplify the general trajectory problem by considering the gravitational force between the spacecraft and only one other body at a time. Even for the case of interplanetary transfer, this simplification will suffice for many calculations. That is, one may consider escape from or capture by a planet to be an interaction between the spacecraft and that particular planet alone, whereas the transfer process is considered an interaction between the spacecraft and the Sun alone. NASA's spectacular multiple flyby missions such as Voyager and Galileo are based on this Keplerian decomposition of the solar system, known as the *patched conic approximation* (or *patched two-body approximation*), discussed in Bate, Mueller, and White [1971].

The strategy of the designers of the Voyager missions was to initially approximate the full N-body solution of the spacecraft's motion as a linkage of several two-body solutions, the well known conic solutions discovered by Kepler. The spacecraft's trajectory as it coasted between two planets was considered as a heliocentric hyperbolic trajectory. The heliocentric trajectory was cleverly chosen to come close to the destination planet, in order to fly by it. When the spacecraft came within the "sphere of

influence"[1] of a planet, it was considered as a hyperbolic conic section trajectory centered on the planet. This patched conic solution could be used as an initial guess for a numerical procedure which produced a fully integrated N-body solution.

High vs. Low Relative Velocities. For missions such as Voyager and Galileo, the speed of the spacecraft relative to the bodies is high and therefore the time during which the acceleration on the spacecraft due to two bodies is comparable is very short, and results in a minor perturbation away from a conic solution. But when one needs to deal with the unpropelled, or ballistic, capture regime of motion,[2] where the relative speed is low, a three-body decomposition of the solar system is necessary.

Some Missions Cannot Be Approximated by the Patched Conic Approach. For Voyager and Galileo, the patched conic approach worked very well. But as space missions have become more demanding, other approaches have become necessary. For example, the Genesis, L_1 Gateway, and multi-moon orbiter trajectories discussed below resemble solutions of the restricted three- and four-body problems much more than two-body problems. In fact, methods based on a patched conic approximation would have a very difficult time finding these complicated trajectories, as they are fundamentally non-Keplerian, restricted N-body solutions.

Taking Better Advantage of N-Body Dynamics. It is possible to satisfy mission constraints using spacecraft solutions which do not take advantage of the N-body dynamics of a system. But this may require using more fuel than is necessary.[3] Worse yet, because of the fuel restrictions on interplanetary spacecraft, some missions may not be possible if only a patched conic approach is used. An interesting example in this category, which also served as motivation for much of our group's work, is the "rescue" of a malfunctioned Japanese space mission to the moon by Belbruno and Miller of JPL in June, 1990. The mission originally had two spacecraft, MUSES-A and MUSES-B; B was to go into orbit around the moon, with A remaining in earth orbit as a communications relay. But B failed and A did not have sufficient fuel to make the journey. However, by utilizing a trajectory concept originally discovered by Belbruno in 1986, which is more energy-efficient than the one planned for B, MUSES-A (renamed Hiten) left Earth orbit in April, 1991 and reached the moon that October. As a

[1] The sphere of influence of a planet is the radius at which the acceleration on a spacecraft due to the planet and the Sun are approximatedly equal (Roy [1988]).

[2] *Ballistic capture* means that no propulsion is necessary (i.e., no ΔV) to achieve a capture orbit at the destination body. In general, this "capture" is temporary.

[3] For example, Dunn [1962] proposed to use a satellite for lunar far side communications by placing it in a position where it would requre approximately 1500 m/s per year for stationkeeping. A few years later, Farquhar [1966] proposed a trajectory for the same mission which used only 100 m/s per year by taking advantage of three-body dynamics.

result, Japan became the third nation to send a spacecraft to the moon. After a series of scientific experiments, Hiten was purposely crashed into the Moon in April, 1993. See Belbruno [2004] for additional details of this fascinating story and a historical perspective on ballistic capture transfers.

SMART-1 is a recent ESA (European Space Agency) mission from the Earth to the Moon (2003–2006). It also uses some of these same ideas; see http://sci.esa.int/science-e/www/area/index.cfm?fareaid=10.

A Hierarchy of Models. We want to make use of the natural dynamics in the solar system as much as possible; that is, we wish to take advantage of the phase space geometry, integrals of motion, and lanes of fast unpropelled travel. We envision generating a trajectory via a hierarchy of models. One starts with simple models which capture essential features of natural dynamics. One then uses simple model solutions as initial guess solutions in more realistic models. The approach described above does this conceptually, using the patched conic approximation to generate the first guess solution. But there are regimes of motion where conics are simply not a good approximation to the motion of the spacecraft. There is much to be gained by starting with not two-body solutions, but three-body solutions to the spacecraft's motion.

The Patched Three-Body Approximation. Motivated by the Belbruno and Miller work, we consider a restricted four-body problem wherein a spacecraft moves under the influence of three massive bodies whose motion is prescribed, as shown schematically in Figure 1.1.1. For Belbruno and Miller, these four bodies were the Sun, the Earth, the Moon and the spacecraft.

To begin with, we restrict the motion of all the bodies to a common plane, so the phase space is only four-dimensional. As in the patched conic approach, the patched three-body approach uses solutions obtained from two three-body problems as an initial guess for a numerical procedure called *differential correction* which converges to a full four-body solution.

As an example of such a problem where there is no control, consider the four-body problem where two adjacent giant planets compete for control of the same comet (e.g., Sun-Jupiter-comet and Sun-Saturn-comet). When close to one of the planets, the comet's motion is dominated by the corresponding planet's perturbation. Between the two planets, the comet's motion is mostly heliocentric and Keplerian, but is precariously poised between two competing three-body dynamics, leading to complicated transfer dynamics between the two adjacent planets.

When we consider a spacecraft with control instead of a comet, we can intelligently exploit the transfer dynamics to construct low energy trajectories with prescribed behaviors, such as transfers between adjacent moons in the Jovian and Saturnian systems (Lo and Ross [1998]). For example, by approximating a spacecraft's motion in the $N + 1$ body gravitational field of Jupiter and N of its planet-sized moons into several segments

of purely three body motion—involving Jupiter, the ith moon, and the spacecraft—we can design a trajectory for the spacecraft which follows a prescribed itinerary in visiting the N moons. In an earlier study of a transfer from Ganymede to Europa, we found our fuel consumption for impulsive burns, as measured by the total norm of velocity displacements, ΔV, to be less than half the Hohmann transfer value (Koon, Lo, Marsden, and Ross [1999]). We found this to be the case for the following example multi-moon orbiter tour shown schematically in Figure 1.2.1: starting beyond

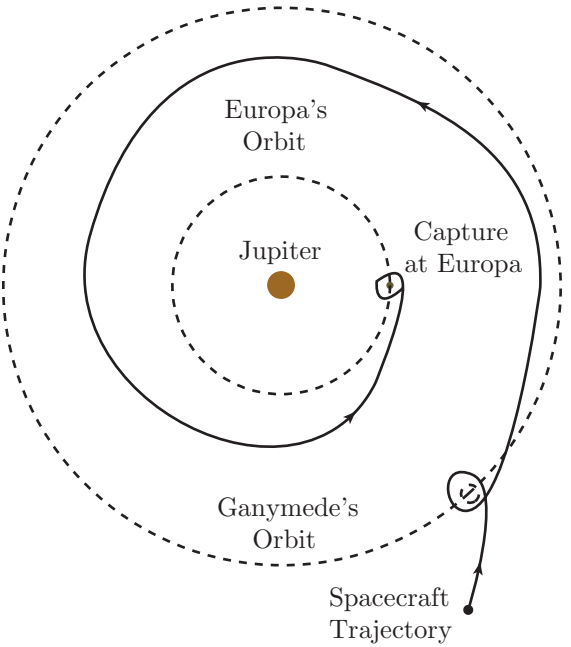


FIGURE 1.2.1. Leap-frogging mission concept: a multi-moon orbiter tour of Jupiter's moons Ganymede and Europa (schematic).

Ganymede's orbit, the spacecraft is ballistically captured by Ganymede, orbits it once, escapes in the direction of Europa, and ends in a ballistic capture at Europa.

One advantage of this multi-moon orbiter approach as compared with the Voyager-type flybys is the "leap-frogging" strategy. In this new approach to mission design, the spacecraft can orbit a moon for a desired number of circuits, escape the moon, and then perform a transfer ΔV to become ballistically captured by a nearby moon for some number of orbits about that moon, etc. Instead of brief flybys lasting only seconds, a scientific spacecraft can orbit several different moons for any desired duration. Furthermore, the total ΔV necessary is *much less than that necessary using purely two-body motion segments*. One can also systematically construct

low energy transfers (i) from the Earth to the Moon using the Sun's gravitational field, and (ii) from lunar libration point orbits to Earth libration point orbits.

Three-Body Dynamics. To patch three-body solutions (the spacecraft's motion in the presence of two bodies), one needs a good handle on what those solutions are. Studying the CR3BP solutions in detail is an interesting topic in its own right. This is a topic that goes back to the basic work of Poincaré in the late 1800s and provided the context in which he developed modern dynamical systems theory and the notion of chaos.

In the CR3BP, we thus have two primaries that move in circles; the smaller third body moves in the gravitational field of the primaries (without affecting them). We typically view the motion in a rotating frame so that the primaries appear stationary. It is important to consider both the planar and the spatial problems, but we shall focus on the planar problem for the moment.

One may derive the equations of motion using a little elementary mechanics as follows. Let the masses of the two primaries be denoted m_1 and m_2 and set $\mu = m_2/(m_1+m_2)$. We can normalize the distance between the primaries to be unity and then in the rotating frame, normalized to rotate with unit angular velocity, the two bodies may be located on the x-axis at the points $(-\mu, 0)$ and $(1-\mu, 0)$. Let the position of the third body be denoted (x, y) in the rotating frame. The kinetic energy of this third body (whose mass we take to be unity) with respect to an inertial frame but written in a frame rotating with unit angular velocity is the usual $\frac{1}{2}mv^2$ expression:

$$K(x, y, \dot{x}, \dot{y}) = \frac{1}{2}\left[(\dot{x} - y)^2 + (\dot{y} + x)^2\right]$$

Let r_1 be the distance from the third body to the first primary; that is, $r_1 = \sqrt{(x+\mu)^2 + y^2}$ and let r_2 be the distance to the second primary, that is, $r_2 = \sqrt{(x-1+\mu)^2 + y^2)}$. Then the gravitational potential energy of the third body is, again in normalized units,

$$U(x, y) = -\frac{1-\mu}{r_1} - \frac{\mu}{r_2}.$$

The Lagrangian of the third body is its kinetic minus potential energies, namely

$$L(x, y, \dot{x}, \dot{y}) = K(x, y, \dot{x}, \dot{y}) - U(x, y).$$

Now one gets the equations of motion simply by writing down the corresponding Euler–Lagrange equations:

$$\ddot{x} - 2\dot{y} = -\frac{\partial \bar{U}}{\partial x}, \qquad \ddot{y} + 2\dot{x} = -\frac{\partial \bar{U}}{\partial y} \tag{1.2.1}$$

where the *effective potential* is

$$\bar{U} = U - \frac{x^2 + y^2}{2}$$

Being Euler–Lagrange equations, there is a conserved energy that one computes via the Legendre transformation to be

$$E = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2\right) + \bar{U}(x, y).$$

Equilibria. These occur when the third body moves in a circular orbit with the same frequency as the primaries, so that it is stationary in the rotating frame. We find these points by finding the equilibrium points, in the standard sense of ODE's, of the equations (1.2.1). It is clear that this task is equivalent to finding the critical points of the effective potential, an analysis that is found in every book on celestial mechanics. The result is that there are five such points. There are three collinear points on the x-axis that were discovered by Euler around 1750 and are denoted L_1, L_2, L_3 and there are two equilateral points discovered by Lagrange around 1760 and are denoted L_4, L_5.[4] They are indicated in Figure 1.2.2.

Equations (1.2.1) may be interpreted as those of a particle moving in an effective potential plus a magnetic field. The graph of the effective potential is shown in Figure 1.2.3. This figure also shows the region one gets by imposing conservation of energy and the simple inequality that the kinetic energy is positive. Thus, at a given energy level E, the third body can only move in the region given by the inequality $E - \bar{U} \geq 0$; this is called the *Hill's region* and is obtained by intersecting the graph of the effective potential with a horizontal plane. An example is shown in the right hand side of Figure 1.2.3 for the Sun-Jupiter-third body system. In this figure, one can see three *realms*, namely the *Sun realm*, the *Jupiter realm* and the *exterior realm* that are connected by the *neck regions*, the left hand neck containing L_1 and the right hand neck containing L_2. For other values of the energy, one or more of these realms may be prohibited due to conservation of energy; that is, the necks may close off.

Of special interest are the two points L_1 and L_2 closest to the secondary body, which a linearized analysis shows are center-saddle points. The famous Lyapunov theorem says that there is a family of periodic orbits surrounding each of these points; one can think of this as meaning that one can "go into orbit about these points". These planar periodic orbits are called *Lyapunov orbits*, while their counterparts in the 3D problem are called *halo* and *Lissajous orbits* (which, by the way involves an interesting bifurcation analysis).

Tubes. In the 3 body problem, a key role is played by the invariant manifolds of these periodic orbits, which we call the *Conley–McGehee tubes*. Also key is a network of homoclinic and heteroclinic orbits connecting these periodic orbits, also discovered in a preliminary way in work of Conley and

[4] Euler [1767] discovered L_1, L_2, and L_3 just a few years before Lagrange [1772] discovered L_4 and L_5, but it is common in the literature to refer to L_1, L_2, L_3 as the "Lagrangian" or "Lagrange points" as well, despite being historically inaccurate.

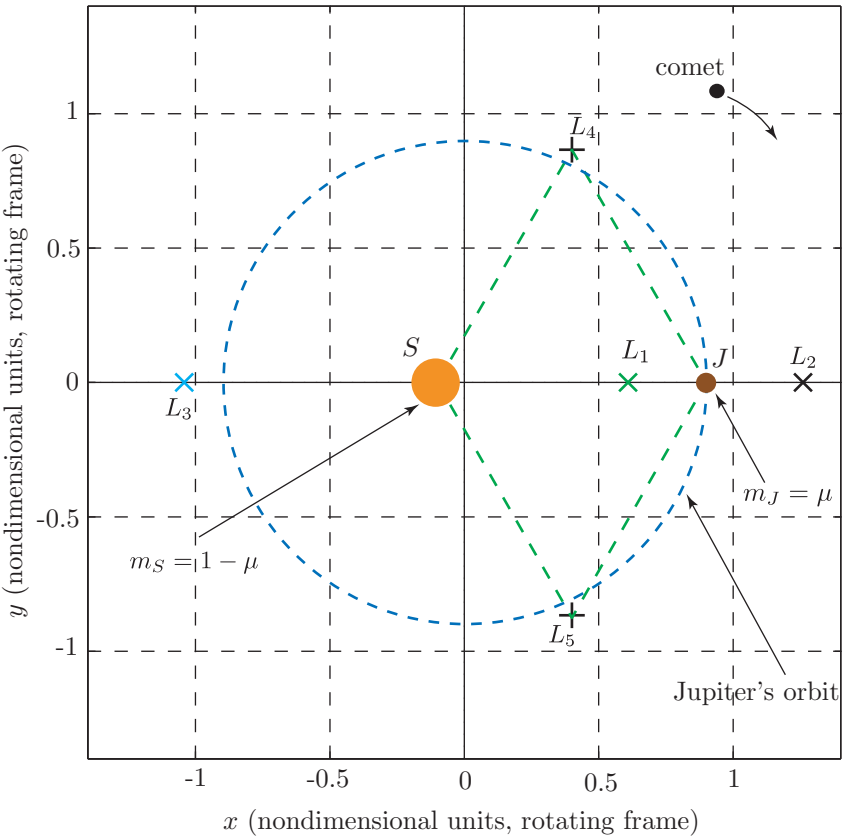


FIGURE 1.2.2. Equilibrium points for the three body problem.

McGehee and was extended and thoroughly investigated in Koon, Lo, Marsden, and Ross [2000]. Some of the reasons that these tubes are important can be seen in the context of specific space missions described below.

In fact, the invariant manifold structures of L_1 and L_2 provide the framework for understanding and categorizing the motions of spacecraft as well as, for example, comets that undergo resonance hopping. Moreover, the stable and unstable invariant manifold tubes associated to periodic orbits around L_1 and L_2 are the phase space conduits transporting material between different realms in a single three body system as well as between primary bodies for separate three-body systems. These tubes can be used to construct new spacecraft trajectories as we will indicate below. It is remarkable that the connecting orbits as well as the associated Conley–McGehee tubes are critical for understanding transport in the solar system as well as in molecular systems. It is quite interesting that some of the same techniques used in the celestial context can also be used in the molecular

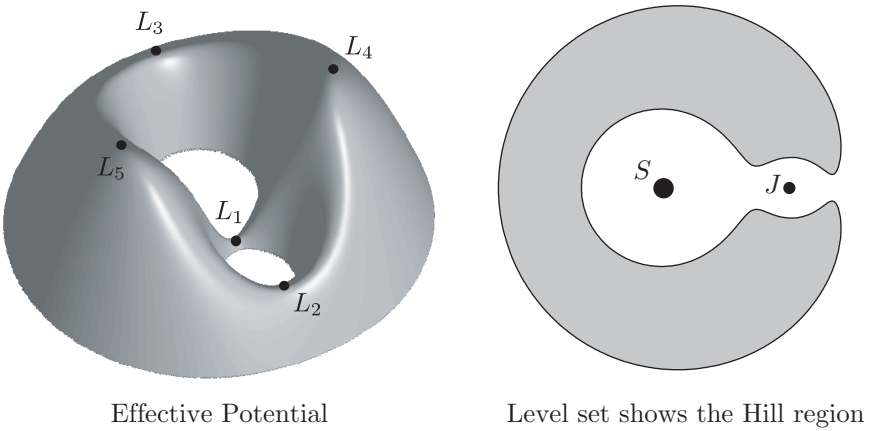


FIGURE 1.2.3. The graph of the effective potential in the 3-body problem. Its critical points are the equilibria.

context, and conversely, techniques from chemistry can be used in celestial problems, as was done by Jaffé, Ross, Lo, Marsden, Farrelly, and Uzer [2002].

Figure 1.2.4 shows some tubes (projected from phase space to configuration space) associated with periodic orbits about L_1, L_2 for the Earth-Moon system. As this figure indicates, it is the tubes that control the capture and escape properties as well as transit and non-transit orbits.

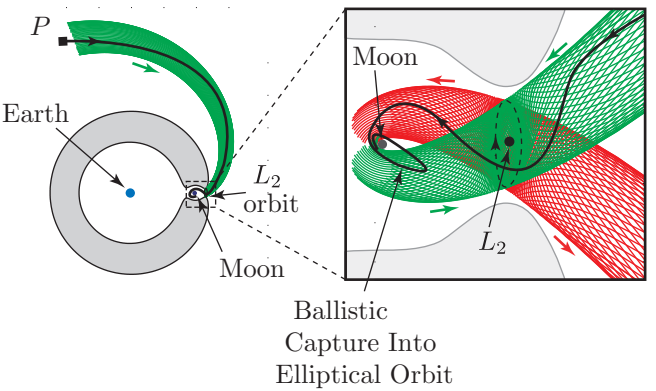


FIGURE 1.2.4. Tube leading to ballistic capture around the Moon (seen in rotating frame).

Supplement: Some Specific Missions

For the complex space missions planned for the near future, greater demands are placed on the trajectory design. In many instances, standard trajectories and classical methods such as the patched two-body approximation are inadequate to support the new mission concepts. Without appropriate and economical trajectories, these missions cannot be achieved. For nearly half a century, space mission planners have depended on trajectory concepts and tools developed in the 1950s and 1960s, based largely on a two-body decomposition of the solar system, the patched conic approach. While that approach remains very valuable for some missions, new trajectory paradigms must be developed to meet today's challenges.

A detailed understanding of the three-body problem, and in particular the dynamics associated with libration points, is absolutely necessary to continue the exploration and development of space.

Figure 1.2.5 shows in metro map format connections between hubs in Earth's neighborhood and beyond. NASA desires to develop a robust and flexible capability to visit several potential destinations. As shown in the figure, NASA has recognized that libration points L_1 and L_2 in the Sun-Earth and Earth-Moon system are important hubs and/or destinations. The fortuitous arrangement of low energy passageways in near-Earth space implies that lunar L_1 and L_2 orbits are connected to orbits around Earth's L_1 or L_2 via low energy pathways.[5] Therefore, a Lunar Gateway Station at the lunar L_1 would be a natural transportation hub to get humanity beyond low-Earth orbit, a stepping stone to the moon, Earth's neighborhood, Mars, the asteroids, and beyond. We will discuss the Lunar L_1 Gateway Station further below.

Because of its unobstructed view of the sun, the Sun-Earth L_1 is a good place to put instruments for doing solar science. NASA's Genesis Discovery Mission has been there, designed completely using invariant manifolds and other tools from dynamical systems theory (Howell, Barden, and Lo [1997]). Other libration point missions include the Solar and Heliospheric Observatory (SOHO),[6] WIND, Advanced Composition Explorer (ACE), WMAP, and ISEE-3/ICE (discussed in Chapter 6).

Genesis Discovery Mission. Launched in August 2001, the Genesis Discovery Mission spacecraft swept up specks of the sun—individual atoms

[5] We will sometimes refer to the Sun-Earth L_1 and L_2 as the Earth's L_1 and L_2, since they are much closer to the Earth than the Sun. Similarly, we will occasionally refer to the Earth-Moon L_1 and L_2 as the lunar or the Moon's L_1 and L_2.

[6] SOHO is a spacecraft mission designed to study the internal structure of the Sun, its extensive outer atmosphere and the origin of the solar wind, the stream of highly ionized gas that blows continuously outward through the solar system. It is a joint project of the European Space Agency (ESA) and NASA. See http://soho.estec.esa.nl for more information.

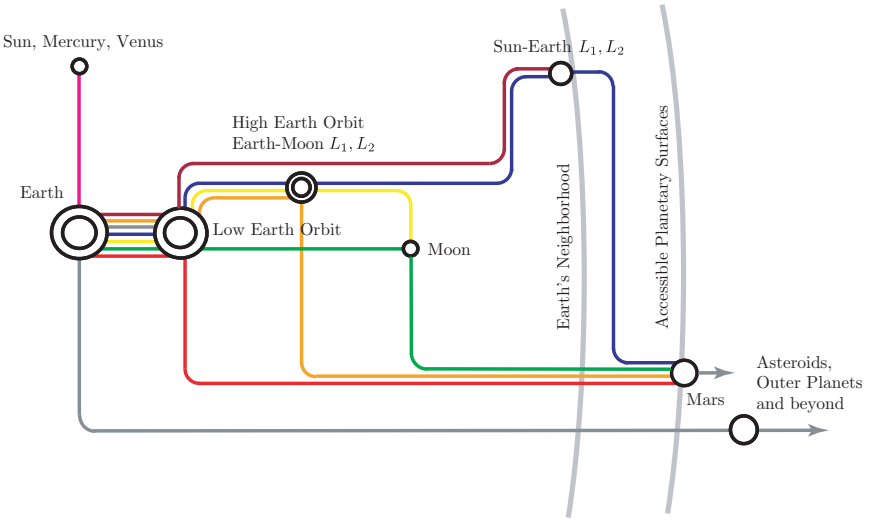


FIGURE 1.2.5. A metro map representation showing hubs connected by low energy passageways in the near-Earth neighborhood and beyond (source: Gary L. Martin, NASA Space Architect).

of the solar wind—on five collector arrays the size of bicycle tires and in an ion concentrator. The goal was to collect solar wind samples and return them safely to the Earth for study into the origins of the solar system. Genesis returned its solar wind cargo to Earth via a sample-return capsule which returned to Earth in September 2004 (see Lo, Williams, Bollman, Han, Hahn, Bell, Hirst, Corwin, Hong, Howell, Barden, and Wilson [2001]).[7] The sample was the first extraterrestrial material brought back to Earth from deep space since the last of the Apollo landings in 1972, and the first to be collected from beyond the moon's orbit.

A reason Genesis was feasible as a mission is that it was designed using low energy passageways. Figure 1.2.6 shows a three-dimensional view of the Genesis trajectory (kindly supplied by Roby Wilson). The spacecraft was launched to a *halo orbit* in the vicinity of the Sun-Earth L_1 and uses a "heteroclinic-like return" in the three-body dynamics to return to Earth.[8]

As noted above, L_1 is the unstable equilibrium point between the Sun and the Earth at roughly 1.5 million km from the Earth in the direction of the Sun. Genesis took a low energy path to its halo orbit, stayed there collecting samples for about 2 years, and returned home on another low

[7] See http://genesismission.jpl.nasa.gov/ for further information.

[8] The orbit is called a "halo orbit" because, as seen from Earth, the flight path follows a halo around the sun. Such orbits were originally named for lunar halo orbits by Farquhar [1968]. Note, setting a spacecraft exactly to the L_1 point is not a good idea, as the spacecraft's radio signals would be lost in the Suns glare.

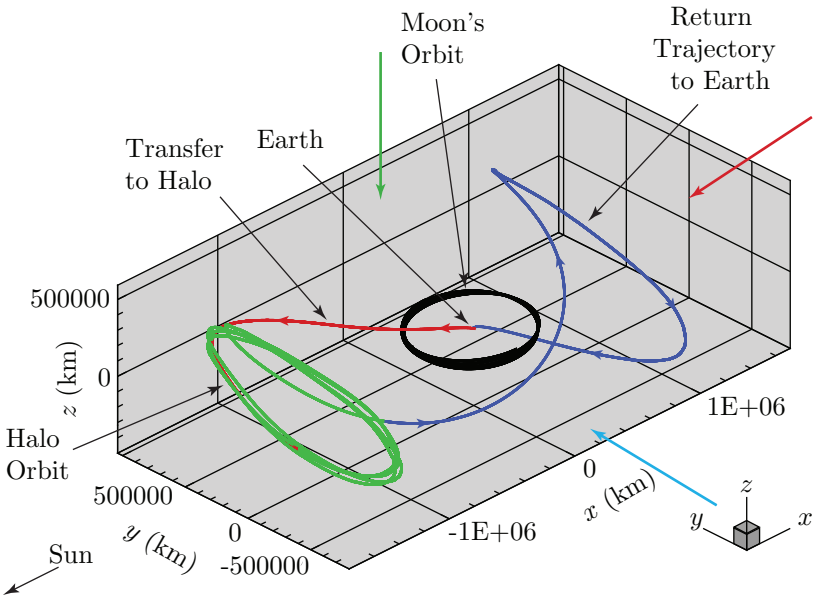


FIGURE 1.2.6. The Genesis Discovery Mission trajectory. The three arrows correspond to the three projections shown in Figure 1.2.7.

energy path.

Figure 1.2.7 shows three orthographic projections of the Genesis trajectory. These figures, plotted in a rotating frame, show the key parts of the trajectory: the transfer to the halo, the halo orbit itself, and the return to Earth. The rotating frame is defined by fixing the x-axis along the Sun-Earth line, the z-axis in the direction normal to the ecliptic, and with the y-axis completing a right-handed coordinate system. The y-amplitude of the Genesis orbit, which extends from the x-axis to the maximum y-value of the orbit, is about 780,000 km (see Figures 1.2.6 and 1.2.7). Note that this is bigger than the radius of the orbit of the Moon, which is about 380,000 km.

As Figures 1.2.6 and 1.2.7 show, the trajectory travels between neighborhoods of L_1 and L_2; L_2 is roughly 1.5 million km on the opposite side of the Earth from the Sun. In dynamical systems theory, this is closely related to the existence of a *heteroclinic connection* between the L_1 and L_2 regions.

The deeper dynamical significance of the heteroclinic connection for the planar three-body problem is that it allows a classification and a construction of orbits using symbolic dynamics, as was shown in Koon, Lo, Marsden, and Ross [2000], and similar phenomena are seen when the third degree of freedom is included, as discussed in Gómez, Koon, Lo, Marsden,

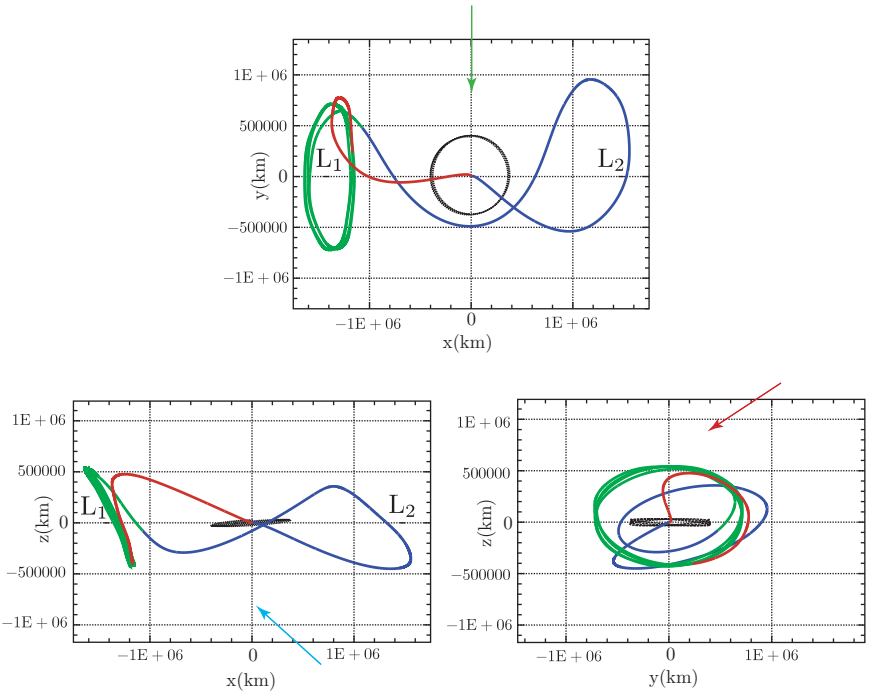


FIGURE 1.2.7. The xy, xz, and yz projections of the three-dimensional Genesis trajectory shown in the preceding figure.

Masdemont, and Ross [2004].

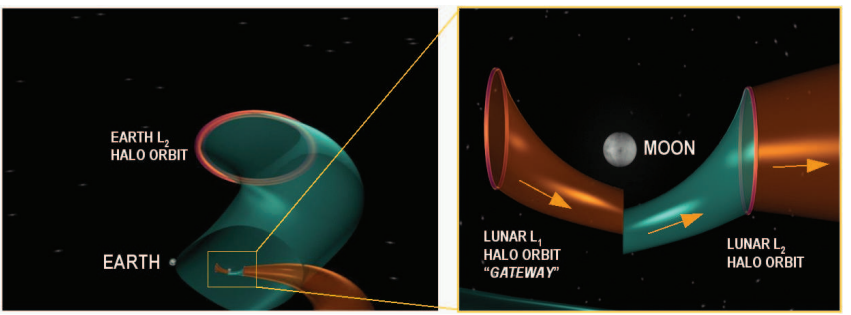
One of the attractive and interesting features of the Genesis trajectory design is that the three year mission, from launch all the way back to Earth return, requires no deterministic maneuver whatsoever and automatically injects into the halo orbit

It is difficult to use traditional classical algorithms[9] to find a near-optimal solution like that of Genesis, so the design of such a low energy trajectory is facilitated by using dynamical systems methods. This is achieved by using the stable and unstable manifolds as guides in determining the end-to-end trajectory. That Genesis performs its huge exotic trajectory using a deterministic ΔV of *zero* (i.e., no fuel) has created a great deal of interest in both the astronautical and mathematical communities.

Lunar L_1 Gateway Station. The work on Genesis has inspired deeper exploration of the dynamics in Earth's neighborhood (see Lo and Ross [2001]). NASA desires to develop a robust and flexible capability to visit

[9]See, for example, Farquhar and Dunham [1981], Farquhar, Muhonen, Newman, and Heuberger [1980], and Farquhar, Muhonen, and Richardson [1977].

several potential destinations, as suggested by the metro map, Figure 1.2.5. A Lunar Gateway Station in the vicinity of the lunar L_1 libration point (between the Earth and the Moon) was proposed as a way station for transfers into the solar system and into the Earth-Sun halo orbits. This is enabled by an historical accident: the energy levels of the Sun-Earth L_1 and L_2 points differ from those of the Earth-Moon system by only 50 m/s (as measured by maneuver velocity). The significance of this coincidence to the development of space cannot be overstated. For example, this implies that the lunar L_1 halo orbits are connected to halo orbits around Earth's L_1 and L_2 via low energy pathways, as illustrated in Figure 1.2.8.



(a) (b)

FIGURE 1.2.8. (a) The fortuitous arrangement of low energy passageways in near-Earth space implies that lunar L_1 and L_2 halo orbits are connected to halo orbits around Earth's L_1 or L_2 via low energy pathways. Many of NASA's future space telescopes located around the Earth's L_1 or L_2 may be built in a lunar L_1 orbit and conveyed to the final destination with minimal fuel requirements. (b) Shown in this close-up are two halo orbits at the lunar L_1 and L_2, respectively, and the set of invariant manifolds that provide the low energy departures from the lunar L_1 orbit.

Many of NASA's future space observatories located around the Earth's L_1 or L_2 may be built in a lunar L_1 orbit and conveyed to the final destination with minimal propulsion requirements. When the spacecraft or instruments require servicing, they may be returned from Earth libration orbits to the lunar L_1 orbit where human servicing may be performed, which was shown to be of vital importance for keeping the Hubble Space Telescope operable. Since a lunar L_1 orbit may be reached from Earth in only three days, the infrastructure and complexity of long-term space travel is greatly mitigated. The same orbit could reach any point on the surface of the Moon within hours, making it a perfect location for the return of humans to the Moon. A lunar L_1 orbit is also an excellent point of departure and arrival for interplanetary flights to Mars, the asteroids, and the outer solar system. Several lunar and Earth encounters may be added to further reduce the launch cost and open up the launch period. A lunar L_1

is therefore a versatile hub for a space transportation system. We discuss the dynamics associated with transfers between Earth and lunar libration points in Chapter 5.

Multi-Moon Orbiters. Using low energy passageways is in no way limited to the inner solar system. For example, consider a spacecraft in the gravity field of Jupiter and its planet-sized moons. A possible new class

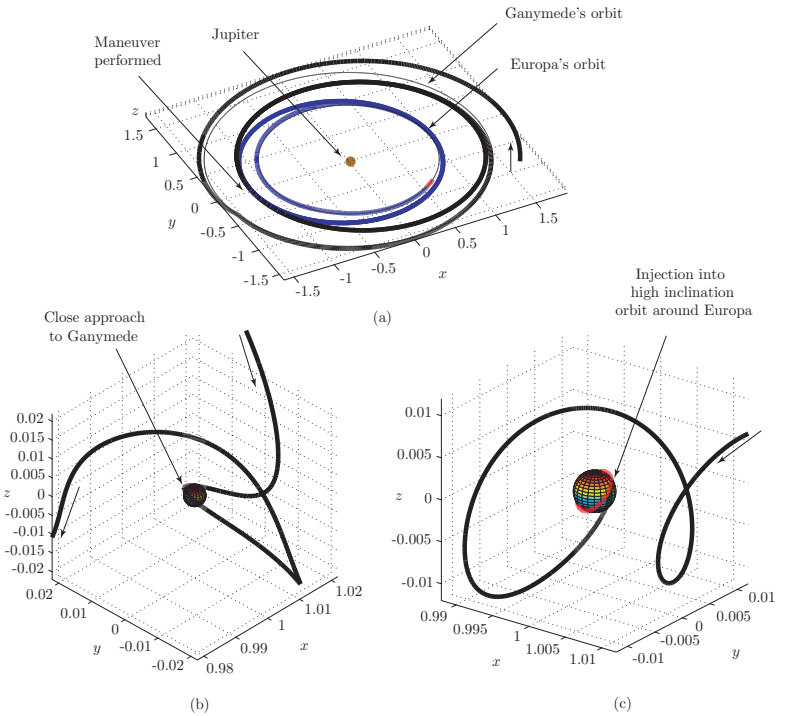


FIGURE 1.2.9. A multi-moon orbiter space mission concept for the Jovian moons. (a) We show a spacecraft trajectory coming into the Jupiter system and transferring from Ganymede to Europa using a single impulsive maneuver, shown in a Jupiter-centered inertial frame. (b) The spacecraft performs one loop around Ganymede, using no propulsion at all, as shown here in the Jupiter-Ganymede rotating frame. (c) The spacecraft arrives in Europa's vicinity at the end of its journey and performs a final propulsion maneuver to get into a high inclination circular orbit around Europa, as shown here in the Jupiter-Europa rotating frame.

of missions to the outer planet moon systems has been proposed by the authors (Koon, Lo, Marsden, and Ross [1999]; Ross, Koon, Lo, and Marsden [2003]). These are missions in which a single spacecraft orbits several moons of Jupiter (or any of the outer planets), allowing long duration observations. Using this multi-moon orbiter approach, a single scientific spacecraft orbits several moons of Jupiter (or any of the outer planets) for

any desired duration, allowing long duration observations instead of flybys lasting only seconds. For example, a multi-moon orbiter could orbit each of the galilean moons—Callisto, Ganymede, Europa, and Io—one after the other, using a technologically feasible amount of fuel. This approach should work well with existing techniques, enhancing trajectory design capabilities for missions such as a future tour to Jupiter's icy moons.

Figure 1.2.9 shows a low energy transfer trajectory from an initial Jovian insertion trajectory to Ganymede. After one orbit around Ganymede including a close approach, the spacecraft heads onward to Europa, ending in a high inclination orbit around the icy moon. The design of such missions is described in Chapter 9.

The low energy trajectory shown in Figure 1.2.9 still requires a significant fuel expenditure. But a very low energy trajectory can be designed using hardly any fuel using the additional technique of *low energy intermoon transfer via resonant gravity assists*. By using very small impulsive thrusts, a spacecraft initially injected into a jovian orbit can be directed into an inclined, elliptical capture orbit around Europa. Enroute, the spacecraft orbits both Callisto and Ganymede for long duration using the ballistic capture and escape methodology described in previous chapters. This example tour is shown in Figure 1.2.10. The design of these kinds of trajectories are discussed in Chapter 10.

1.3 Organization of the Book

In Chapter 2, we develop some basic terminology for the restricted three-body problem (R3BP) and describe the local dynamics near saddle-center equilibrium points, building on the work of Conley [1968]. In Chapters 2 through 5, we look at the *planar, circular* R3BP (i.e., the PCR3BP), where a particle moves in the field of two massive bodies on circular orbits about their common center of mass and is restricted to move in their plane of motion. The PCR3BP has five equilibria, $L_i, i = 1, \ldots, 5$. We focus our attention on two of these equilibrium points, L_1 and L_2. Periodic orbits about these points are shown to exist. Furthermore, stable and unstable manifolds of these periodic orbits in phase space with a $S^1 \times \mathbb{R}^1$ (cylindrical) geometry are shown to exist. The periodic orbits considered reside in bottleneck regions of the energy manifold, separating large zones, otherwise known as realms, associated with motion about one mass, the other mass, or both masses. The cylinders have the physical property that all motion through the bottleneck in which the periodic orbit resides must occur through the interior of these surfaces. The cylinders thus mediate the global transport of test particles between large zones of the energy surface which are separated by the bottlenecks.

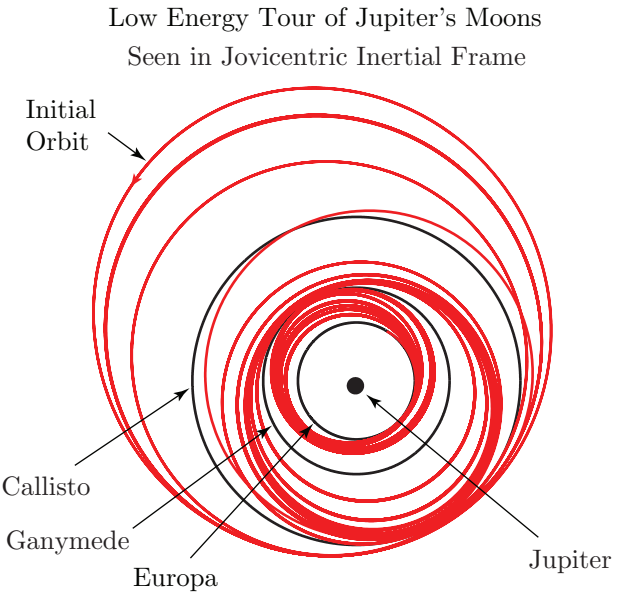


FIGURE 1.2.10. **The multi-moon orbiter mission concept for the jovian moons** involves long duration orbits of Callisto, Ganymede, and Europa, allowing for extensive observation. By utilizing resonant gravity assists with the moons, in addition to ballistic capture and escape orbits leading toward or away from temporary capture orbits about a moon, a tour can be constructed using very little fuel. The trajectory shown is a simulation of a restricted 5-body problem and requires a ΔV of only 22 m/s. The multi-moon orbiter is a general concept applicable for any multi-moon system and is not limited to the specific example shown.

In Chapter 3, the local picture is extended to larger regions of phase space. The cylindrical stable and unstable invariant manifolds of L_1 and L_2, referred to as *tubes*, are shown to play a crucial role in our understanding of the global connectivity of the phase space. In particular, the existence of a heteroclinic connection between pairs of periodic orbits is numerically demonstrated, one around L_1 and the other around L_2, with the two periodic orbits having the same energy. This heteroclinic connection, along with previously known homoclinic connections, allows us to prove a theorem on the global orbit structure of the PCR3BP. This theorem can also be taken as a proof of "horseshoe-like" chaos in the system.

In Chapter 4, we use the tubes and their intersections to compute orbits of desired itineraries with respect to the aforementioned realms of the phase space and make the connection to the design of space mission trajectories.

In Chapter 5, we construct solutions of the restricted four-body problem using solutions of the restricted three-body problem as building blocks. This approach, which is appropriate for some low energy space mission trajectories, is the patched three-body approximation. We demonstrate the

approach with the numerical construction of a low energy Earth-to-Moon trajectory which uses the Sun's gravitational field. We also look at transfers from a Lunar L_1 Gateway Station to a libration orbit around the Earth-Sun L_2 point.

Earlier chapters focus on the two degree-of-freedom problem, but in Chapter 6 we begin to discuss the three degree-of-freedom circular restricted three-body problem (CR3BP). With the addition of a third degree of freedom, new phenomena emerge. In particular, the orbital taxonomy near the libration points gets more complicated. Actual missions require three-dimensional capabilities, such as control of the latitude and longitude of a spacecraft's escape from and entry into a planetary or moon orbit. For example, a future mission to send a probe to orbit Europa may desire a capture into a high inclination polar orbit around Europa (Sweetser, Maddock, Johannesen, Bell, Penzo, Wolf, Williams, Matousek, and Weinstein [1997]; Ludwinski, Guman, Johannesen, Mitchell, and Staehle [1998]; Scheeres, Guman, and Villac [2001]; Villac and Scheeres [2001]). Three-dimensional capability is also required when decomposing an N-body system into three-body subsystems which are not co-planar, such as the Earth-Sun-spacecraft and Earth-moon-spacecraft systems. (The tilt in the orbital planes of the Earth around the sun and the moon around the Earth is about 5 degrees.) These demands necessitate dropping the restriction to planar motion, and extension of earlier results to the spatial model.

In particular, Chapter 6 introduces halo orbits and how to compute them. Halo orbits are large three dimensional orbits shaped like the edges of a potato chip. The computation of halo orbits follows standard nonlinear trajectory computation algorithms based on parallel shooting. Due to the sensitivity of the problem, an accurate first guess is essential, since the halo orbit is actually an unstable orbit (albeit with a fairly long time constant in the Sun-Earth system, about 180 days). This first guess is provided by a high order analytic expansion using the Lindstedt–Poincaré method.

In Chapter 7, we discuss the numerical computation of the end-to-end trajectory for the Genesis mission. We will discuss the computation of invariant manifolds using various methods, e.g., Floquet theory, and how to find solutions in more realistic (ephemeris) models using differential correction.

In Chapter 8, we discuss optimal control, trajectory correction maneuvers, and station keeping for three-body spacecraft trajectories. When one computes spacecraft trajectories, several idealistic assumptions are made. When a ΔV is computed for an impulsive burn trajectory, the calculation assumes that the rocket can instantaneously and flawlessly deliver the exact amount ΔV necessary to impart the boost in velocity. In practice, however, errors are introduced by the rocket engine and other sources which must be corrected by subsequent maneuvers.

In Chapter 9, we show that the invariant manifold structures of the collinear libration points in the three degree-of-freedom problem still act

as the separatrices between two types of motion (as they do in the two degree-of-freedom problem): (i) inside the invariant manifold tubes, the motion consists of transit through a neck, a set of paths called transit orbits; (ii) outside the tubes, no such transit motion is possible. We design an algorithm for constructing orbits with any prescribed itinerary and obtain some initial results for a basic itinerary. Furthermore, we apply these new techniques to the construction of a three-dimensional multi-moon orbiter tour of the Jovian moon system. By approximating the dynamics of the Jupiter-Europa-Ganymede-spacecraft four-body problem as two three-body subproblems, we seek intersections (in position space only) between the tubes of transit orbits enclosed by the stable and unstable manifold tubes. We design an example low energy transfer trajectory from an initial Jovian insertion trajectory, leading to Ganymede and finally to Europa, ending in a high inclination orbit around Europa.

Besides providing a full description of different kinds of libration motions in a large vicinity of these points, in Chapter 9 we numerically demonstrate the existence of heteroclinic connections between pairs of libration orbits, one around the libration point L_1 and the other around L_2. Since these connections are asymptotic orbits, no maneuver is needed to perform the transfer from one libration point orbit to the other. The knowledge of these orbits may provide the backbone for many interesting orbits in the future. The numerical demonstration is achieved using normal form methods which we describe at the end of the chapter.

In Chapter 10, we discuss advanced topics related to spacecraft trajectory design using multi-body effects. The incorporation of resonant gravity assists to further reduce ΔV is discussed, from the point of view of lobe dynamics in the restricted three-body problem. The design of very low energy multi-moon orbiters is discussed. What is found is that there is an inverse relationship between time-of-flight and fuel consumption. The longer the trajectory, the more the cumulative effect of multi-body dynamics and the lower the fuel needed. We also discuss the incorporation of low thrust into the framework described in this book, by which we mean the design of low thrust control laws which best take advantage of natural multi-body effects.

2

Motion near the Collinear Equilibrium Points

2.1 Introduction

In Chapter 1 we motivated the study of the gravitational interaction of three bodies from the field of spacecraft trajectory design. In this chapter and the next, we begin the detailed analysis of a simplification of the general three-body problem, one in which we study the motion of a test particle with negligible mass compared to the other two.

Suppose the two massive bodies move in circular orbits about their common center of mass and the mass of the third body is too small to affect the motion of the two more massive bodies. The problem of the motion of the third body is called the circular, restricted, three-body problem, henceforth referred to as the CR3BP. If we further restrict the motion of the third body to be in the orbital plane of the other two bodies, the problem is called the ***planar circular restricted three-body problem***, or the PCR3BP.

At first glance this problem may seem to have little application to motion in the solar system. After all, the observed orbits of solar system objects are non-circular, albeit with small eccentricities. However, the hierarchy of orbits and masses in the solar system (e.g., sun, planet, satellite, ring particle) means that the CR3BP provides a good approximation for certain systems, especially the *qualitative* behavior of those systems. In fact, we

need consider only the lower dimensional PCR3BP to understand a wide range of dynamical behaviors for the third body.

This chapter is roughly divided into two parts. In the first part (§2.2-2.5), we describe the equations for the problem with particular reference to a constant of the motion, the Hamiltonian energy. We demonstrate the relationship between curves defined by the Hamiltonian energy and the orbital path of the particle. We discuss the location and stability of equilibrium points.

In the second part (§2.6-2.9), we consider the motion of particles near two important equilibria, the libration points L_1 and L_2, which will be explained shortly. By considering the motion near these two equilibria in detail, we lay the foundation for understanding the global picture of the PCR3BP phase space, to be covered in Chapter 3.

We will be focusing on particular aspects of the three-body problem which are important for the discussion in later chapters. The books by Szebehely [1967] and Marchal [1990] provide authoritative coverage of the literature on the subject. For other general introductions to the three-body problem, see Abraham and Marsden [2008], Meyer and Hall [1992], Holmes [1990], or Simó [1999].

2.2 Planar Circular Restricted Three-Body Problem

Problem Description. Consider the motion of a ***particle*** P of negligible mass moving under the gravitational influence of two masses m_1 and m_2, referred to as the primary masses, or simply the ***primaries***. In some cases, we refer to m_1 as the primary and m_2 as the ***secondary***. Assume that m_1 and m_2 have circular orbits about their common center of mass. The particle P is free to move in the plane defined by the circular orbits of the primaries, but cannot affect their motion.

In the context of this chapter, we can imagine that m_1 represents the sun and m_2 represents a planet, and we are concerned with the motion of P, a comet or spacecraft of much smaller mass.

The system is made nondimensional by the following choice of units: the unit of mass is taken to be $m_1 + m_2$; the unit of length is chosen to be the constant separation between m_1 and m_2 (e.g., the distance between the centers of the sun and planet); the unit of time is chosen such that the orbital period of m_1 and m_2 about their center of mass is 2π. The universal constant of gravitation then becomes $G = 1$. It then follows that the common mean motion, n, of the primaries is also unity. We will refer to this system of units as ***nondimensional*** or ***normalized*** units throughout the book.

We will use the normalized units for nearly all the discussions in this book. When appropriate, we can convert to dimensional units (e.g., km, km/s, seconds) to scale a problem. This will be particularly important in Chapter 5 when we will decompose a N-body problem into several 3-body subproblems. The conversion from units of distance, velocity, and time in the unprimed, normalized system to the primed, dimensionalized system is

$$\begin{array}{ll} \text{distance} & d' = Ld, \\ \text{velocity} & s' = Vs, \\ \text{time} & t' = \frac{T}{2\pi}t, \end{array}$$

where L is the distance between the centers of m_1 and m_2, V is the orbital velocity of m_1, T is the orbital period of m_1 and m_2.

The only parameter of the system is the ***mass parameter***,

$$\mu = \frac{m_2}{m_1 + m_2}.$$

If we assume that $m_1 > m_2$, then the masses of m_1 and m_2 in this system of units are, respectively,

$$\mu_1 = 1 - \mu \quad \text{and} \quad \mu_2 = \mu,$$

where $\mu \in [0, \frac{1}{2}]$, and thus $\mu_1 \geq \mu_2$. The phase space of the system is highly dependent on the mass parameter. A table of mass parameters and dimensional values L (in km), V (in km/s), and T (in seconds), for several pairs of masses in the solar system is provided below in Table 2.2.1.

System	μ	L	V	T
Sun-Jupiter	9.537×10^{-4}	7.784×10^{8}	13.102	3.733×10^{8}
Sun-(Earth+Moon)	3.036×10^{-6}	1.496×10^{8}	29.784	3.147×10^{7}
Earth-Moon	1.215×10^{-2}	3.850×10^{5}	1.025	2.361×10^{6}
Mars-Phobos	1.667×10^{-8}	9.380×10^{3}	2.144	2.749×10^{4}
Jupiter-Io	4.704×10^{-5}	4.218×10^{5}	17.390	1.524×10^{5}
Jupiter-Europa	2.528×10^{-5}	6.711×10^{5}	13.780	3.060×10^{5}
Jupiter-Ganymede	7.804×10^{-5}	1.070×10^{6}	10.909	6.165×10^{5}
Jupiter-Callisto	5.667×10^{-5}	1.883×10^{6}	8.226	1.438×10^{6}
Saturn-Mimas	6.723×10^{-8}	1.856×10^{5}	14.367	8.117×10^{4}
Saturn-Titan	2.366×10^{-4}	1.222×10^{6}	5.588	1.374×10^{6}
Neptune-Triton	2.089×10^{-4}	3.548×10^{5}	4.402	5.064×10^{5}
Pluto-Charon	1.097×10^{-1}	1.941×10^{4}	0.222	5.503×10^{5}

TABLE 2.2.1. **Table of m_1-m_2 systems in the solar system.** Source: The first three are the values used in Koon, Lo, Marsden, and Ross [2000, 2001b]. The others are from the Jet Propulsion Laboratory's solar system dynamics website: `http://ssd.jpl.nasa.gov/`.

Aside: When is the PCR3BP an Appropriate Starting Model?

For astronomical phenemena like resonance transition of Jupiter comets, the PCR3BP is considered an adequate starting model (see Koon, Lo, Marsden, and Ross [2000]). The comets of interest are mostly heliocentric, i.e., their motion is dominated by the sun's gravitational force, but their perturbations from purely two-body motion about the sun are dominated by Jupiter's gravitation. Their motion is nearly in Jupiter's orbital plane, and the small eccentricity of Jupiter's orbit (i.e., it is nearly circular) plays little role during resonant transition. In a more detailed study, and for verification that one can really believe the PCR3BP model, one need to take into account Jupiter's eccentricity, the effect of three-dimensionality of the motion, and, for instance, the perturbation from Saturn.

For the design of the spacecraft trajectory for the Genesis Discovery Mission, knowledge of heteroclinic behavior provided the necessary insight in searching for the desired solution. The Genesis trajectory is an example of a trajectory which takes into account the gravity field of multiple bodies, in this case the Earth, moon, and sun, which at times have comparable and competing effects on the spacecraft's motion. Other examples of trajectories in N-body gravitational fields will be considered in this book, including a trajectory to orbit several of Jupiter's planet-sized moons. As we will see, the study of PCR3BP provides a systematic method for the numerical construction of initial pieces from which the final trajectory in the N-body field can be generated.

2.3 Equations of Motion

In this section, we familiarize the reader with some of the terminology of the PCR3BP and the all important concept of viewing the motion in the rotating frame. However, for the benefit of later chapters, the equations of motion will be derived for the general problem, the CR3BP. The planar version can be obtained simply by restricting the motion of the particle to the orbital plane of the primaries.

There are several ways to derive the equations of motion for P in the field of m_1 and m_2. We will go over a few of the ways, emphasizing the Hamiltonian structure for this system.[1] A simple technique is to use the

[1] For example, Whittaker [1927] and Abraham and Marsden [2008] use time dependent canonical transformation theory to transform the problem from an inertial frame to a rotating frame.

covariance of the Lagrangian formulation and use the Lagrangian written using coordinates in a moving frame, as given in Marsden and Ratiu [1999]. This method directly gives the equations in Lagrangian form and the associated Hamiltonian form is given by the Legendre transformation. We shall discuss this approach later in this section, but we begin with a derivation starting with Newton's equations, $\mathbf{F} = \frac{d}{dt}(m\mathbf{v})$, where m and $\mathbf{v}$ are the mass and velocity of the particle and $\mathbf{F}$ is the sum of external forces on the particle. If the particle's mass is constant in time, this reduces to the familiar $\mathbf{F} = m\mathbf{a}$, where $\mathbf{a} = \dot{\mathbf{v}}$

Transformation between the Inertial and Rotating Frames. Let X-Y-Z be an inertial frame with origin at the m_1-m_2 center of mass, as in Figure 2.3.1, where the X-Y plane is the orbital plane of the primaries. Consider the set of axes x and y depicted in Figure 2.3.1. The x-axis lies along the line from m_1 to m_2 with the y-axis perpendicular to it, completing a right-handed coordinate system. The x-y frame rotates

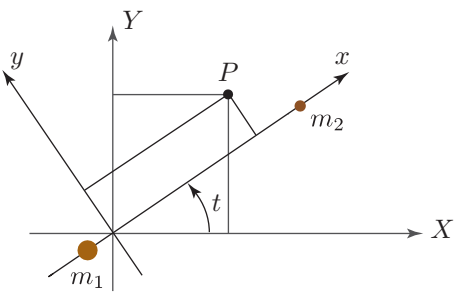


FIGURE 2.3.1. **Inertial and rotating frames.** The rotating coordinate system with coordinates x and y moves counterclockwise with unit angular velocity relative to the inertial frame with coordinates X and Y. The z-axis (which coincides with the Z-axis) is pointing out of the plane and is not shown here.

with respect to the X-Y inertial frame with an angular velocity equal to the mean motion, n, of either mass (unity in the normalized units). We will refer to this coordinate frame throughout the book as the ***rotating frame*** or the m_1-m_2 ***rotating frame***. Assume that the two frames coincide at $t = 0$. Let (X, Y, Z) and (x, y, z) be the position of P in the inertial and rotating frames, respectively. In normalized units, we have the following transformation of the particle's position between the two frames:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = A_t \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{2.3.1}$$

where

$$A_t = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.3.2}$$

Differentiating gives us the transformation of velocity components from the rotating to the inertial frame:

$$\begin{aligned} \begin{pmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{pmatrix} &= \dot{A}_t \begin{pmatrix} x \\ y \\ z \end{pmatrix} + A_t \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix}, \\ &= -A_t J \begin{pmatrix} x \\ y \\ z \end{pmatrix} + A_t \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix}, \\ &= A_t \begin{pmatrix} \dot{x} - y \\ \dot{y} + x \\ \dot{z} \end{pmatrix}, \end{aligned} \tag{2.3.3}$$

where

$$J = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Rotating Frame. The rotating frame is shown in Figure 2.3.2. The larger mass, m_1, is located at $(-\mu_2, 0, 0)$ and the smaller mass, m_2, at $(\mu_1, 0, 0)$. This is also true in the inertial frame when $t = 0$. At general times t,

$$\begin{aligned} (X_1, Y_1, Z_1) &= (-\mu_2 \cos t, -\mu_2 \sin t, 0), \\ (X_2, Y_2, Z_2) &= (\mu_1 \cos t, \mu_1 \sin t, 0), \end{aligned}$$

are the inertial frame positions of m_1 and m_2, respectively.

The rotating system of coordinates takes some getting used to, especially for those comfortable with conic section orbits in the inertial frame from the two-body problem. We emphasize that in Figure 2.3.2, (x, y, z) are the position coordinates of P relative to the positions of the m_1 and m_2, *not relative to an inertial frame.*

Aside: Why Use the Rotating Frame?

It is quite common to see much more structure in carefully chosen rotating frames than in stationary inertial frames, even in cases where there is no "obvious" preferred rotating frame. This example of *pattern evocation*

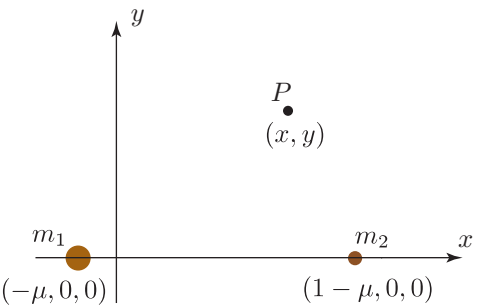


FIGURE 2.3.2. **Rotating coordinate frame.** The planar circular restricted three-body problem as viewed, not in an inertial frame, but in the rotating frame, where m_1 and m_2 are at fixed positions along the x-axis. As with the previous figure, the z-axis is not shown.

is well illustrated by the double spherical pendulum.[2]

For example, in the comet resonance problem considered in Koon, Lo, Marsden, and Ross [2000], the inertial frame, Figure 2.3.3(a), gives little insight into the dynamics of the comet resonance transition. However, when the comet undergoing a resonance transition has its motion viewed in the sun-Jupiter rotating frame, one can clearly observes more structure. For example, comets come near to the libration points, L_1 and L_2, to pass from beyond Jupiter's orbit to within Jupiter's orbit, getting temporarily gravitationally captured by the planet in the process.

As seen in Figures 2.3.3(b) and (c), one can see how closely the orbit of comet Oterma follows the plots of invariant manifolds related to L_1 and L_2 in the position space of the rotating frame. Analytically, in this preferred sun-Jupiter rotating frame, the equations of motion of the comet are time-independent. The system of equations have an integral of motion and equilibrium points which allow us to bring in all the tools of dynamical system theory.

Gravitational Potential. The gravitational potential which the particle experiences due to m_1 and m_2 (in normalized units) is

$$\mathcal{U} = -\frac{\mu_1}{r_1} - \frac{\mu_2}{r_2} - \frac{1}{2}\mu_1\mu_2 \tag{2.3.4}$$

[2] Marsden and Scheurle [1995]

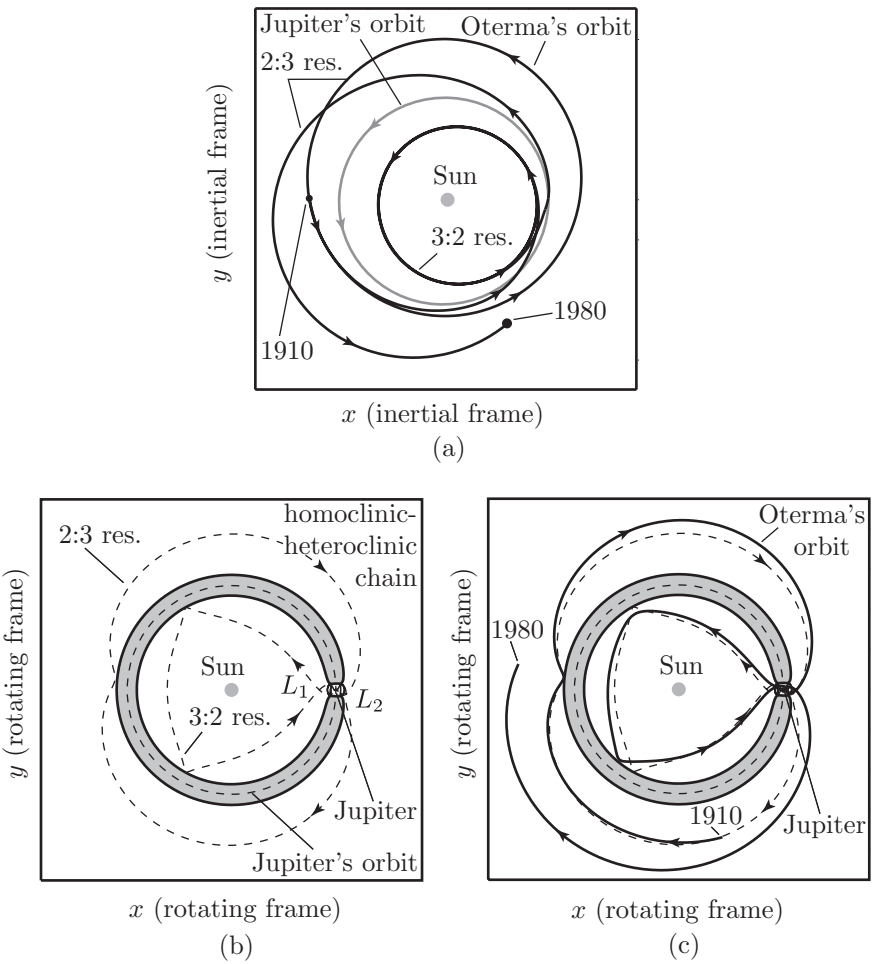


FIGURE 2.3.3. (a) Orbit of comet Oterma in sun-centered inertial frame during time interval AD 1910–1980 (ecliptic projection). (b) Some manifolds associated to L_1 and L_2 for the energy of Oterma in the planar, circular, restricted three-body problem, as seen in the rotating frame with the sun and Jupiter fixed. (c) The orbit of Oterma, transformed into the rotating frame, follows closely the invariant manifolds associated to L_1 and L_2. Distances are in Astronomical Units (AU).

where r_1 and r_2 are the distances of P from m_1 and m_2, respectively, given by

$$r_1^2 = (X + \mu_2 \cos t)^2 + (Y + \mu_2 \sin t)^2 + Z^2,$$
$$r_2^2 = (X - \mu_1 \cos t)^2 + (Y - \mu_1 \sin t)^2 + Z^2.$$

The constant last term in the expression for $\mathcal{V}$ is added by convention (see, e.g., Llibre, Martinez, and Simó [1985]), and will not affect the equations

of motion.

Newtonian Approach: Inertial Frame. In the inertial frame, the Newtonian equations of motion are

$$\ddot{X} = -\mathcal{U}_X, \qquad \ddot{Y} = -\mathcal{U}_Y, \qquad \ddot{Z} = -\mathcal{U}_Z, \tag{2.3.5}$$

where $\mathcal{U}_X$, $\mathcal{U}_Y$, and $\mathcal{U}_Z$ are the partial derivatives of $\mathcal{U}$ with respect to X, Y, Z respectively. This system is time-dependent. One can now make a transformation of variables to the variables (x, y, z) by direct computation.[3] This procedure leads to the same equations of motion in terms of (x, y, z) as the methods below.

Lagrangian Approach: Inertial Frame. Consider the Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0, \tag{2.3.6}$$

where the mechanical system is described by generalized coordinates $(q^1, \ldots, q^n)$. One usually chooses the Lagrangian L to be of the form kinetic minus potential energy. See Marsden and Ratiu [1999] or other books on mechanics for a discussion.

In the inertial frame, the Lagrangian $\mathcal{L}$ is kinetic minus potential energies and is given by

$$\mathcal{L}(X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}, t) = \frac{1}{2}(\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2) - \mathcal{U}(X, Y, Z, t).$$

Lagrangian Approach: Rotating Frame. In the rotating frame, the Lagrangian L is given by

$$L(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \frac{1}{2}\left((\dot{x} - y)^2 + (\dot{y} + x)^2 + \dot{z}^2\right) - U(x, y, z).$$

The Lagrangian is now *time-independent*, simplifying the analysis of solutions.

We obtain this formula for L by simply rewriting the kinetic and potential energy of the inertial frame Lagrangian $\mathcal{L}$ in rotating coordinates. From Eq. (2.3.3), the kinetic energy is

$$\frac{1}{2}(\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2) = \frac{1}{2}\left((\dot{x} - y)^2 + (\dot{y} + x)^2 + \dot{z}^2\right).$$

Also, since both the distances r_1 and r_2 are invariant under rotation, the gravitational potential is

$$U(x, y, z) = -\frac{\mu_1}{r_1} - \frac{\mu_2}{r_2} - \frac{1}{2}\mu_1\mu_2, \tag{2.3.7}$$

[3] See Marsden and Ratiu [1999] for this type of calculation.

where r_1 and r_2 are expressed in rotating coordinates as

$$r_1^2 = (x+\mu_2)^2 + y^2 + z^2,$$
$$r_2^2 = (x-\mu_1)^2 + y^2 + z^2.$$

The theory of moving systems says that one can simply write down the Euler–Lagrange equations in the rotating frame and one will get the correct equations. It is a very efficient generic method for computing equations for either moving systems or for systems seen from moving frames. See Marsden and Ratiu [1999] for more information.

In the present case, the Euler–Lagrange equations are given by

$$\frac{d}{dt}(\dot{x} - y) = \dot{y} + x - U_x,$$
$$\frac{d}{dt}(\dot{y} + x) = -(\dot{x} - y) - U_y.$$
$$\frac{d}{dt}\dot{z} = -U_z.$$

After simplification, we have

$$\begin{aligned} \ddot{x} - 2\dot{y} &= -\bar{U}_x, \\ \ddot{y} + 2\dot{x} &= -\bar{U}_y, \\ \ddot{z} &= -\bar{U}_z, \end{aligned} \tag{2.3.8}$$

where

$$\begin{aligned} \bar{U}(x,y) &= -\frac{1}{2}(x^2+y^2) + U(x,y,z), \\ &= -\frac{1}{2}(x^2+y^2) - \frac{\mu_1}{r_1} - \frac{\mu_2}{r_2} - \frac{1}{2}\mu_1\mu_2, \qquad (2.3.9) \\ &= -\frac{1}{2}(\mu_1 r_1^2 + \mu_2 r_2^2) - \frac{\mu_1}{r_1} - \frac{\mu_2}{r_2}, \qquad (2.3.10) \end{aligned}$$

is the ***augmented*** or ***effective potential*** and the subscripts in (2.3.8) denote its partial derivatives. The planar version of (2.3.8) can be obtained by setting $z = \dot{z} = 0$. This form of the equations has been studied in detail in Szebehely [1967].

Hamiltonian Approach: Rotating Frame. Whenever one has a Lagrangian system, one can transform it to Hamiltonian form by means of the Legendre transformation:

$$p_i = \frac{\partial L}{\partial \dot{q}^i}; \qquad H(q^i, p_i) = \sum_{i=1}^{n} p_i \dot{q}^i - L(q^i, p_i),$$

to get the equations in Hamiltonian form

$$\dot{q}^i = \frac{\partial H}{\partial p_i}; \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i}.$$

In our case, the Legendre transformation is given by

$$\begin{aligned} p_x &= \frac{\partial L}{\partial \dot{x}} = \dot{x} - y, \\ p_y &= \frac{\partial L}{\partial \dot{y}} = \dot{y} + x, \\ p_z &= \frac{\partial L}{\partial \dot{z}} = \dot{z}, \end{aligned}$$

and so we obtain the Hamiltonian function

$$\begin{aligned} H(x, y, z, p_x, p_y, p_z) &= p_x \dot{x} + p_y \dot{y} + p_z \dot{z} - L \\ &= \frac{1}{2}((p_x + y)^2 + (p_y - x)^2 + p_z^2) + \bar{U}, \end{aligned} \tag{2.3.11}$$

where p_x, p_y, and p_z are the momenta conjugate to x, y, and z, respectively.

Hence the Hamiltonian equations are given by

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p_x} = p_x + y, \\ \dot{y} &= \frac{\partial H}{\partial p_y} = p_y - x, \\ \dot{z} &= \frac{\partial H}{\partial p_z} = p_z, \\ \dot{p}_x &= -\frac{\partial H}{\partial x} = p_y - x - \bar{U}_x, \\ \dot{p}_y &= -\frac{\partial H}{\partial y} = -p_x - y - \bar{U}_y, \\ \dot{p}_z &= -\frac{\partial H}{\partial z} = -\bar{U}_z. \end{aligned} \tag{2.3.12}$$

One can also transform from the inertial frame to the rotating frame by using the theory of canonical transformations. This method, while the one classically used, is more complicated. See Whittaker [1927] for details.

Notice that both the Lagrangian and the Hamiltonian form of the equations in rotating coordinates (x, y, z) give a time-independent system. Viewed as a dynamical system, it is a dynamical system in a six-dimensional ***phase space***, viewed as either $(x, y, z, \dot{x}, \dot{y}, \dot{z})$ or (x, y, z, p_x, p_y, p_z) space, subsets of $\mathbb{R}^6$ which exclude the singularities at the positions of the primaries.

Energy Integral and Jacobi Constant. Since the equations of motion of the CR3BP (2.3.12) are Hamiltonian and independent of time, they have an ***energy integral*** of motion. We use the symbol H when we regard the energy as a function of positions and momenta, as in (2.3.11), and E when we regard it as a function of the positions and velocities,

$$E(x, y, \dot{x}, \dot{y}, \dot{z}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + \bar{U}(x, y, z) \tag{2.3.13}$$

Physically, the measurement of the particle's position and velocity in either the inertial or rotating frames determines the value of the energy associated with the particle's motion.

The celestial mechanics and dynamical astronomy communities uses $-2E$, which is called the ***Jacobi integral*** and is given by

$$C(x, y, z, \dot{x}, \dot{y}, \dot{z}) = -(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - 2\bar{U}. \tag{2.3.14}$$

Usually in those communities, the existence of the Jacobi integral is derived directly from the equations of motion. The computation is straightforward:

$$\begin{aligned}\frac{d}{dt}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) &= 2(\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z}) \\ &= 2[\dot{x}(2\dot{y} - \bar{U}_x) + \dot{y}(-2\dot{x} - \bar{U}_y) + \dot{z}(-\bar{U}_z)] = 2\frac{d}{dt}(-\bar{U}),\end{aligned}$$

so we get

$$\frac{d}{dt}C = \frac{d}{dt}\left(-(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - 2\bar{U}\right) = 0.$$

Throughout the book, we will use the terms "energy," "energy integral," "Jacobi integral," and "Jacobi constant" to refer to the same concept—the most important integral determining the motion of the particle. As they differ in sign, we will make it clear from the context when we are referring to increasing energy (decreasing Jacobi constant), etc. In general, there are no other integrals constraining the motion of the particle, making the PCR3BP a *non-integrable* problem.

2.4 Energy Surface and Realms of Possible Motion

In the two-body Kepler problem, one may divide the phase space into two major categories, based on the values of the Keplerian energy, $E_{Kep} = -\frac{1}{2a}$, where a is the semimajor axis of the test particle's orbit around the central massive body. The following two cases divide the phase space into two major categories of possible motion for the test particle.

- (i) $E_{Kep} < 0$: Negative Keplerian energies correspond to bound motion of the test particle about the single massive body, i.e., elliptical and circular orbits.

- (ii) $E_{Kep} > 0$: Positive Keplerian energies correspond to unbound motion, i.e., hyperbolic orbits coming from and going to infinity.

The critical case of zero energy orbits between these two are the unbound parabolic orbits. If we restrict ourselves to the *planar* Kepler problem, we have a four-dimensional phase space, which we can view as an open set in $\mathbb{R}^4$: two position coordinates and their two corresponding velocities. For each real value, e, the equation, $E_{Kep} = e$, describes a three-dimensional set in the four-dimensional phase space, termed the *energy surface corresponding to energy* e. The phase space can be viewed as a many layered "onion," each layer or leaf corresponding to a value of the energy. One says that the energy surfaces foliate the phase space.

In the three-body problem, the picture is more complicated, but we can follow a similar strategy of categorizing the possible motion of the test particle by energy, this time the three-body energy given in (2.3.13). In the next few chapters, we will concentrate on the study of the PCR3BP with $z = \dot{z} = 0$.

Energy Surface. Let $\mathcal{M}$ be the ***energy manifold*** or ***energy surface*** given by setting the energy integral (2.3.13) equal to a constant, i.e.,

$$\mathcal{M}(\mu, e) = \{(x, y, \dot{x}, \dot{y}) \mid E(x, y, \dot{x}, \dot{y}) = e\}, \tag{2.4.1}$$

where e is a constant. For a fixed μ and energy e, one can consider the surface $\mathcal{M}(\mu, e)$ as a three-dimensional surface embedded in the four-dimensional phase space.

Hill's Region: the Region of Possible Motion. The projection of this surface onto position space in the rotating frame, the x-y plane, is the region of possible motion for a particle of energy e in the field of two masses with mass parameter μ. Let $M(\mu, e)$ denote this projection,[4]

$$M(\mu, e) = \{(x, y) \mid \bar{U}(x, y) \leq e\}, \tag{2.4.2}$$

known historically as the ***Hill's region***. The boundary of $M(\mu, e)$ is known as the ***zero velocity curve***, and plays an important role in placing bounds on the motion of the particle.

Zero Velocity Curves: the Boundaries of the Hill's Region. The zero velocity curves are the locus of points in the x-y plane where the kinetic energy, and hence the velocity, $v = \sqrt{\dot{x}^2 + \dot{y}^2}$, vanishes, i.e., $\frac{1}{2}v^2(x, y) = e - \bar{U}(x, y) = 0$. From (2.4.2), it is clear that the particle is only able to move on the side of this curve for which the kinetic energy is positive. The other side of the curve, where the kinetic energy is negative and motion is not possible, is known as the *forbidden realm*.

Recall that the energy E is given by (2.3.13). Fixing the energy function to be a constant, i.e., $E(x, y, \dot{x}, \dot{y}) = e$, is like fixing a height in the plot of

[4]Note that our convention is to use script letters for a region in the energy surface (including the energy surface itself, $\mathcal{M}$) and italicized capital letters for that region's projection onto the position space (e.g., M).

the effective potential, $\bar{U}(x, y)$, given in Figure 2.4.1. Consider the surface of the effective potential in Figure 2.4.1 and note the following features.

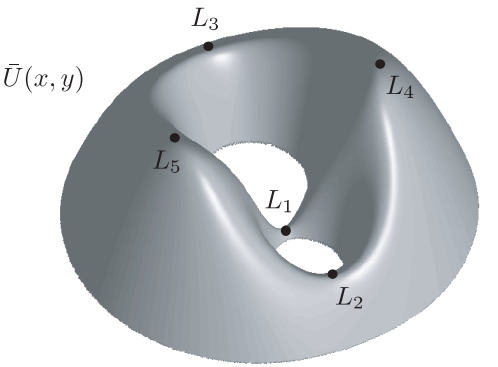


FIGURE 2.4.1. The plot of the effective potential $\bar{U}(x, y)$ for $\mu = 0.3$. The critical points are at the locations of the five equilibrium points, $L_i, i = 1, \ldots, 5$.

- Near either m_1 or m_2, we have a potential well.
- Far away from either m_1 or m_2, the term that corresponds to the centrifugal force dominates $\bar{U}$ in (2.3.9), i.e., $||\frac{1}{2}(x^2 + y^2)||/||\frac{1-\mu}{r_1} + \frac{\mu}{r_2}|| \gg 1$, and we have another potential well.
- By multivariable calculus, one finds that there are five critical points where the slope is zero: three saddle points along the x axis and two symmetric points off the x axis. As will be covered in the next section, these points are the x-y locations of the equilibrium points for a particle in the rotating frame, i.e., a particle placed here at rest with respect to m_1 and m_2 (zero initial velocity), will stay at rest for all time (zero acceleration). We label these points $L_i, i = 1, \ldots, 5$, as in Figure 2.4.1.
- Let E_i be the energy of a particle at rest at L_i, then $E_5 = E_4 > E_3 > E_2 > E_1$. Thus, L_1 is the location of the lowest energy equilibrium point and L_4 and L_5 are the highest energy equilibrium points. Since the energy is measured in a rotating frame, we cannot determine the stability properties of all the equilibrium points from their ordering by energy (e.g., L_4 and L_5 are spectrally stable for small μ, despite being energy maxima, as covered in Szebehely [1967]).

The Five Cases of the Hill's Region. For a given μ there are five basic configurations for the Hill's region, corresponding to five intervals of energy value, e, in (2.4.1). We refer to these basic configurations as ***energy cases***, or simply ***cases***. The cases are shown in Figure 2.4.2. We will show

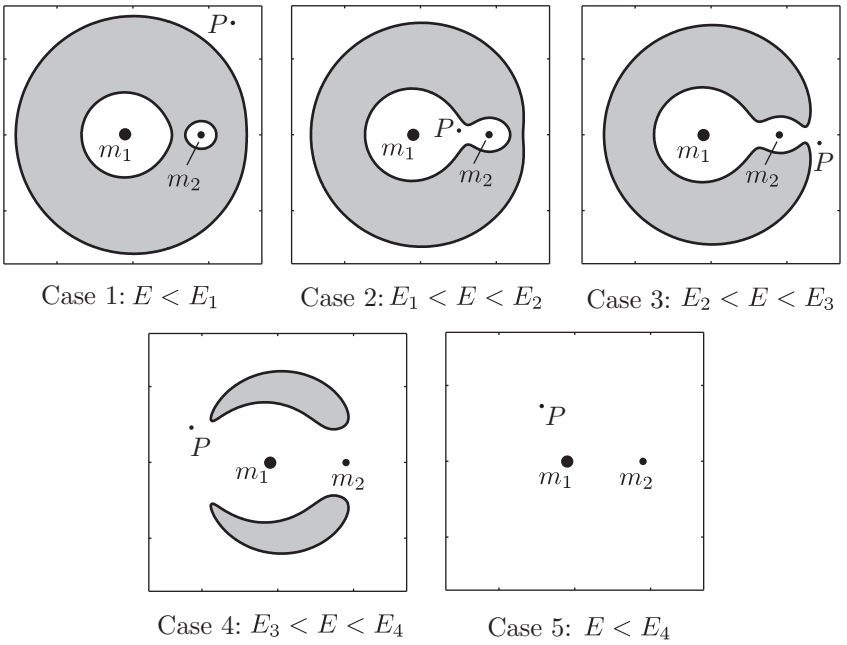


FIGURE 2.4.2. **Realms of possible motion.** Zero velocity curves for five values of the energy, one in each of the cases as described in the text, are shown on the x-y plane for $\mu = 0.3$. These curves bound the zone, in white, accessible by the particle, P, for a given energy value, $E = e$. The part of the x-y plane which is shaded is inaccessible for a given energy, and known as the *forbidden realm*. The outermost accessible realm, known as the *exterior realm*, extends to infinity. In the fifth case (e), the forbidden realm vanishes and motion over the entire x-y plane is possible.

how to compute the energy intervals corresponding to these cases.

Contour plots of the effective potential give the five cases of Hill's region. The white areas in Figure 2.4.2 are the Hill's region and the shaded areas are the forbidden realm.

- **Case 1,** $E < E_1$: If the energy of the particle is below E_1, the particle cannot move between the realms around m_1 and m_2.
- **Case 2,** $E_1 < E < E_2$: If the energy is just above E_1, a "neck" between the realms around m_1 and m_2 opens up, permitting the particle to move between the two realms. The L_1 point is in this neck. We will see in §2.6 that the transport between the two adjacent realms is controlled by invariant manifold structures associated to L_1. The particle is still barred from moving between these two realms and the exterior realm extending to infinity.
- **Case 3,** $E_2 < E < E_3$: This is the case that concerns us the most; when the energy is just above E_2. The particle can move between the

vicinity of m_1 and m_2 and the exterior realm via a neck around L_2.

- **Case 4,** $E_3 < E < -\frac{3}{2} = E_4 = E_5$: In this case the energy is above E_3 but below that of E_4 and E_5, which is always $-\frac{3}{2}$. The particle can pass directly from the vicinity of m_1 to the exterior realm via a neck around L_3.
- **Case 5,** $-\frac{3}{2} < E$: If the energy is above $E_4 = E_5 = -\frac{3}{2}$, the forbidden realm disappears. Case 5 is where the particle is free to move in the entire x-y plane.

Realms of Possible Motion. A glance at Figure 2.4.2 reveals that, beginning in case 1, there are three main ***realms of possible motion***, or simply ***realms***. Considering, for example, Figure 2.4.2(a), the large region surrounding m_1 is referred to as the m_1 ***realm***, sometimes referred to as the ***interior realm***. The small region surrounding m_2 is the m_2 ***realm***. The realm which lies outside both the m_1 and m_2 realms, and extends to infinity, is the ***exterior realm***. For case 1, the realms are separated. Moving up in energy to case 2, a neck around L_1 opens up between the m_1 and m_2 realms, permitting the particle to pass between the two. An additional neck opens up around L_2 when we move up in energy to case 3, permitting travel between all three realms. Our main interest in this book will be case 3; but for comparison we shall occasionally bring up case 2.

The critical values of E which separate these five cases are the values $E_i, i = 1, \ldots, 4$ previously mentioned, corresponding to the equilibrium points $L_i, i = 1, \ldots, 4$. These values can be easily calculated for small μ as will be shown in the following section. The graphs of the E_i as a function of μ are shown in Figure 2.4.3. For instance, for case 3 the energy value lies between E_2 and E_3 which are the energy values of the libration points L_2 and L_3, respectively.

2.5 Location of the Equilibrium Points

Written in first-order form, the equations of motion for the PCR3BP are

$$\begin{aligned} \dot{x} &= v_x, \\ \dot{y} &= v_y, \\ \dot{v}_x &= 2v_y - \bar{U}_x, \\ \dot{v}_y &= -2v_x - \bar{U}_y. \end{aligned} \tag{2.5.1}$$

To find equilibrium points, we set the right-hand sides of the system equal to zero. We see that equlibria in (x, y, v_x, v_y) space are of the form $(x_e, y_e, 0, 0)$, where (x_e, y_e) are critical points of the effective potential function $\bar{U}(x, y)$ shown in Figure 2.4.1.

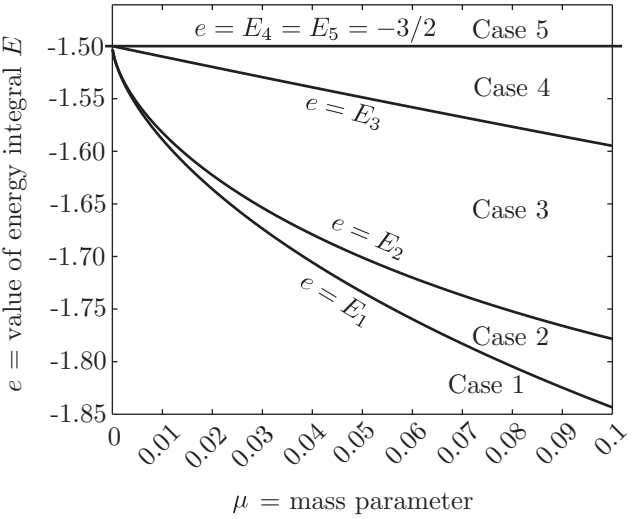


FIGURE 2.4.3. **The graphs of the E_i as a function of μ** partition the μ-e plane into the five cases of possible motion. The Hill's regions for cases 1 through 4 are shown in Figure 2.4.2.

As described in the previous section, the PCR3BP admits five equilibrium point solutions, which are shown in Figure 2.5.1:

- three collinear equilibria on the x-axis, called L_1, L_2, L_3; and
- two equilateral points called L_4, L_5.

These equilibria can be found as follows.

The Equilateral Points. First, we seek solutions that do not lie on the line joining the primaries, i.e., $y \neq 0$. Using the distances r_1, r_2 as variables and the relation

$$x^2 + y^2 = (1-\mu)r_1^2 + \mu r_2^2 - \mu(1-\mu),$$

we see that $\bar{U}$ can be written as

$$-\bar{U}(r_1, r_2) = \frac{1}{2}(1-\mu)r_1^2 + \frac{1}{2}\mu r_2^2 + \frac{1-\mu}{r_1} + \frac{\mu}{r_2}.$$

Using the chain rule, it is straightforward to show that if $y \neq 0$, then $\bar{U}(r_1, r_2)$ and $\bar{U}(x, y)$ have the same critical points.

$$\bar{U}_x = \bar{U}_{r_1}\frac{\partial r_1}{\partial x} + \bar{U}_{r_2}\frac{\partial r_2}{\partial x} = \bar{U}_{r_1}\frac{x+\mu}{r_1} + \bar{U}_{r_2}\frac{x-(1-\mu)}{r_2} = 0$$
$$\bar{U}_y = \bar{U}_{r_1}\frac{\partial r_1}{\partial y} + \bar{U}_{r_2}\frac{\partial r_2}{\partial y} = \bar{U}_{r_1}\frac{y}{r_1} + \bar{U}_{r_2}\frac{y}{r_2} = 0$$

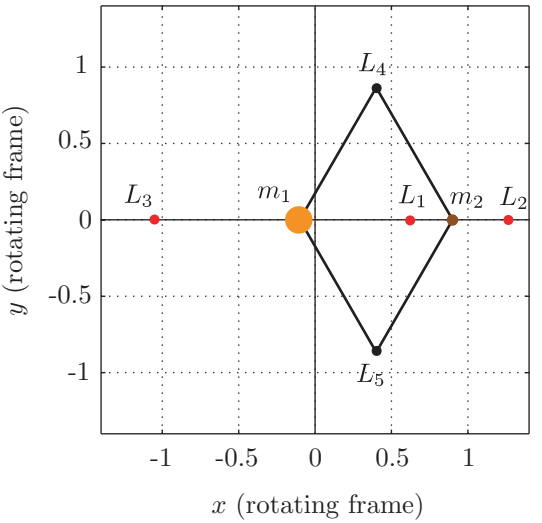


FIGURE 2.5.1. **Equilibrium points of the circular restricted three-body problem** in the x-y plane of the frame rotating with the mean motion of the orbit of m_1 and m_2. A particle placed at rest at such a point will remain at rest for all time. The points marked with an 'x' are linearly unstable. Those marked with a '+' are unstable for $\mu \geq \mu_0 \simeq 0.038521$ and spectrally stable otherwise (see Szebehely [1967] for details). The points shown here are for $\mu = 0.3$.

Solving the following systems

$$0 = -\bar{U}_{r_1} = \mu r_2 - \frac{\mu}{r_2^2},$$
$$0 = -\bar{U}_{r_2} = (1-\mu) r_1 - \frac{(1-\mu)}{r_2^2},$$

we get the unique solution $r_1 = r_2 = 1$.

This solution lies at the vertex of an equilateral triangle whose base is the line segment joing the two primaries. By convention, the one in the upper half-plane is denoted L_4, and the one in the lower half-plane is denoted L_5. These equilibria are 60° ahead of and behind m_2 in its orbit about the $m_1 - m_2$ center of mass, respectively.

The Collinear Points. Consider equilibria along the line of primaries where $y = 0$. In this case the effective potential function has the form

$$\bar{U}(x,0) = -\frac{1}{2}x^2 - \frac{1-\mu}{|x+\mu|} - \frac{\mu}{|x-1+\mu|}.$$

It can be determined that $\bar{U}(x,0)$ has precisely one critical point in each of the following three intervals along the x-axis: (i) $(-\infty, -\mu)$, (ii) $(-\mu, 1-\mu)$ and (iii) $(1-\mu, \infty)$.

This is because $\bar{U}(x,0) \to -\infty$ as $x \to \pm\infty$, as $x \to -\mu$, or as $x \to 1-\mu$. So $\bar{U}$ has at least one critical point on each of these three intervals. Also,

$$\frac{d^2\bar{U}}{dx^2}(x,0) = -1 - \frac{1-\mu}{|x+\mu|^3} - \frac{\mu}{|x-1+\mu|^3},$$

is always negative, so $\bar{U}(x,0)$ is concave in each of the intervals. Therefore, $\bar{U}(x,0)$ has precisely one critical point in each of these three intervals. A sketch of the graph of $\bar{U}(x,0)$ is given in Figure 2.5.2.

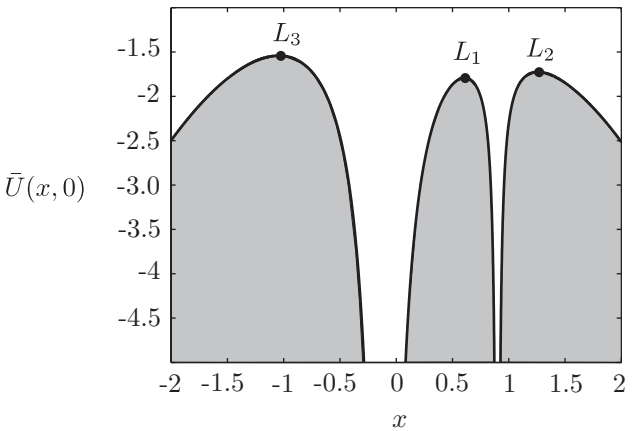


FIGURE 2.5.2. **The graph of** $\bar{U}(x,0)$ for $\mu = 0.1$ is shown. The solid line is the intersection of $\bar{U}(x,y)$ in Figure 2.4.1 with the plane defined by $y = 0$. At the x locations of m_1 and m_2, the function plunges to $-\infty$. The maxima of $\bar{U}(x,0)$ correspond to the unstable collinear equilibrium points L_1, L_2, and L_3.

Locating the Collinear Equilibria. Computation of the x values of the collinear points requires finding the maxima of the function $\bar{U}(x,0)$, i.e., the solutions of $\frac{d}{dx}\bar{U}(x,0) = 0$ which is a quintic equation after simplification. The distance from $L_i, i = 1,2$ to the smaller primary is given by the unique positive solution γ_i of the following equation:

$$\gamma^5 \mp (3-\mu)\gamma^4 + (3-2\mu)\gamma^3 - \mu\gamma^2 \pm 2\mu\gamma - \mu = 0 \tag{2.5.2}$$

where the upper sign is for γ_1 and the lower one for γ_2 (see Szebehely [1967]). A similar equation can be found for γ_3, the distance between L_3 to the larger primary.

Historically, a lot of effort has been spent finding the series expansion for such solutions. Here, we will write down two of those, from Szebehely

[1967], that are most useful for us:

$$\gamma_1 = r_h(1 - \frac{1}{3}r_h - \frac{1}{9}r_h^2 + \ldots), \tag{2.5.3}$$

$$\gamma_2 = r_h(1 + \frac{1}{3}r_h - \frac{1}{9}r_h^2 + \ldots), \tag{2.5.4}$$

where $r_h = (\frac{\mu}{3})^{\frac{1}{3}}$, the *Hill radius*, is the radius of the Hill sphere in the spatial problem. The Hill sphere is the 'bubble' in 3-D position space surrounding m_2 inside of which the gravitational field of m_2 and m_1 may have a comparable effect on the particle's motion.

Locating these points to a higher accuracy numerically is straightforward. $\gamma_i, i = 1, 2$ can be solved by the Newton method, using r_h or the above series expansion as an initial solution for the quintic equation (2.5.2).

As an example calculation, consider the motion of a particle in the Sun-Jupiter system ($\mu = 9.537 \times 10^{-4}$ from Table 2.2.1). The Hill radius is $r_h = 6.825 \times 10^{-2}$, and $\gamma_1 = 6.666 \times 10^{-2}$ to third-order in r_h via Eq. (2.5.3). Solving the quintic equation (2.5.2) numerically yields $\gamma_1 = 6.668 \times 10^{-2}$, and thus the x position of L_1 is $x_{L_1} = x_{m_1} - \gamma_1 = (1 - \mu) - \gamma_1 = 0.9324$.

A Note on Terminology. Throughout the literature covering the equilibrium points in the PCR3BP, the points are given various names, such as libration points, Lagrange points, and Lagrangian points. In this book, we will restrict ourselves to the terms equilibrium or libration point.

2.6 Linearization near the Collinear Equilibria

In this section, we begin the study of the behavior of particle trajectories near the two libration points L_1 and L_2, which on either side of m_2 along the x-axis. As will become clear, we are particularly interested in particles which have an energy just above that of the critical point L_2, that is, $E > E_2$ in case 3.

As shown in Figure 2.4.2(c), the region of possible motion for case 3 contains a neck *about each libration point*. Thus, a particle starting in the exterior realm may pass through the neck around L_2 to the m_2 realm, and subsequently pass through the neck around L_1 to the m_1 realm, and so on.

The aim in the next few sections is to describe the geometry of trajectories in the neck regions. We begin by considering the equations of motion linearized near the equilibrium point inside the neck region. By virtue of Moser's generalization of a theorem of Lyapunov *all the qualitative results of such a discussion carry over to the full nonlinear equations* (see Moser [1958]). See the Supplement below for more details on this theorem.

In the following sections, we will use L to denote either L_1 or L_2. Furthermore, for a fixed energy E, consider a neighborhood of L in the energy surface, whose position space projections are the neck regions described previously. We refer to this neighborhood as the ***equilibrium region*** and denote it by $\mathcal{R}$ on the energy surface. Following our adopted convention, $\mathcal{R}$ has the position space projection R.

Hamiltonian Approach. To find the linearized equations around the collinear libration point L with coordinates $(x_e, y_e, \dot{x}_e, \dot{y}_e) = (x_e, 0, 0, 0)$, we need the quadratic terms of the Hamiltonian H in equation (2.3.11) as expanded about $(x_e, y_e, p_{xe}, p_{ye}) = (x_e, 0, 0, x_e)$. After making a coordinate change with $(x_e, 0, 0, x_e)$ as the origin, these quadratic terms form the Hamiltonian function for the linearized equations, which we shall call H_l.

$$H_l = \frac{1}{2}\left[(p_x + y)^2 + (p_y - x)^2 - ax^2 + by^2\right], \tag{2.6.1}$$

where a and b are defined by $a = 2\bar{\mu} + 1$, and $b = \bar{\mu} - 1$ and where

$$\bar{\mu} = \mu|x_e - 1 + \mu|^{-3} + (1 - \mu)|x_e + \mu|^{-3}. \tag{2.6.2}$$

It can be shown that both a and b are positive constants.

A short computation gives the linearized equations in the Hamiltonian form

$$\begin{aligned} \dot{x} &= \frac{\partial H_l}{\partial p_x} = p_x + y, \\ \dot{y} &= \frac{\partial H_l}{\partial p_y} = p_y - x, \\ \dot{p}_x &= -\frac{\partial H_l}{\partial x} = p_y - x + ax, \\ \dot{p}_y &= -\frac{\partial H_l}{\partial y} = -p_x - y - by. \end{aligned} \tag{2.6.3}$$

Lagrangian Approach. Using the inverse Legendre transformation: $v_x = p_x + y, v_y = p_y - x$, where v_x, v_y correspond to velocity in the rotating coordinate system, we obtain the linearized equations in the Lagrangian form

$$\begin{aligned} \dot{x} &= v_x, \\ \dot{y} &= v_y, \\ \dot{v}_x &= 2v_y + ax, \\ \dot{v}_y &= -2v_x - by, \end{aligned} \tag{2.6.4}$$

which is the linearization of the equations (2.5.1) around the equilibrium point $(x_e, 0, 0, 0)$.

The integral H_l of (2.6.1) now appears as

$$E_l = \frac{1}{2}(v_x^2 + v_y^2 - ax^2 + by^2), \tag{2.6.5}$$

which corresponds to the energy integral E of the restricted problem. Notice that the zero-surface of the integral E_l corresponds to the energy surface which passes through the libration point. We shall therefore study solutions of equations (2.6.4) on the surface $E_l = \varepsilon > 0$ which corresponds to the case where the Hill's region contains a neck about the libration point.

Supplement: Moser's Theorem and Local Integrals near the Collinear Equilibrium Points

Suppose we have a time-independent, analytic Hamiltonian system of differential equations with two degrees of freedom. Suppose these equations have a non-degenerate equilibrium point with one pair of real and one pair of imaginary eigenvalues, $\pm\lambda$ and $\pm i\nu$. We can assume, without loss of generality, that the phase space coordinates (x_1, x_2, y_1, y_2) are chosen so that the Hamiltonian function assumes the following form:

$$H(x, y) = \lambda x_1 y_1 + \tfrac{1}{2}\nu(x_2^2 + y_2^2) + O_3(x, y),$$

where $x = (x_1, x_2), y = (y_1, y_2)$ and the symbol $O_n(\cdot, \cdot)$ denotes terms of order n or higher in the variables displayed.

In particular, the equilibrium point has coordinates $x = y = 0$ and the differential equations are obtained from H as

$$\begin{aligned}
\dot{x}_1 &= H_{y_1} = \lambda x_1 + O_2(x, y),\\
\dot{y}_1 &= -H_{x_1} = -\lambda y_1 + O_2(x, y),\\
\dot{x}_2 &= H_{y_2} = \nu y_2 + O_2(x, y),\\
\dot{y}_2 &= -H_{y_1} = \nu x_2 + O_2(x, y).
\end{aligned}$$

The linearized equations are similarly obtained from a Hamiltonian function which consists of the quadratic terms of H or, equivalently, by dropping the terms of order two or higher in the above equations. Solutions of these linearized equations are conveniently written as

$$\begin{aligned}
x_1(t) &= x_1^0 e^{\lambda t}, \qquad & y_1(t) &= y_1^0 e^{-\lambda t},\\
z(t) &= x_2(t) + i y_2(t) = z^0 e^{-\nu t}, & &
\end{aligned}$$

where the constants x_1^0, y_1^0 and $z^0 = x_2^0 + i y_2^0$ are the initial conditions.

These linearized equations admit integrals in addition to the Hamiltonian function; namely, the functions $x_1 y_1$ and $|z|^2 = x_2^2 + y_2^2$ are both constant along solutions. A special case of a theorem by Moser [1958] states that the full non-linear equations admit "local" integrals analogous to these: thus there are two power series in x and y beginning respectively with quadratic

terms x_1y_1 and $x_2^2+y_2^2$ which converge in some neighborhood of $x=y=0$ and such that the corresponding functions are constants along pieces of solutions lying in the domain of convergence.

A special case of Moser's theorem is stated by Conley [1969] in a form suited to the PCR3BP. In this statement ξ and η are real variables and ζ is complex.

Theorem (Moser). Let $x=y=0$ correspond to a critical point as decribed above. Then there exists a (real) analytic, transformation

$$\begin{aligned} x_1 &= \xi + O_2(\xi,\eta,\zeta,\bar{\zeta}), & y_1 &= \eta + O_2(\xi,\eta,\zeta,\bar{\zeta}),\\ z &= x_2 + iy_2 = \zeta + O_2(\xi,\eta,\zeta,\bar{\zeta}), \end{aligned}$$

as well as power series α and β in the variables $\chi=\xi\eta$ and $|\zeta|^2$ of the form

$$\begin{aligned} \alpha &= \lambda + O_1(\chi,|\zeta|^2),\\ \beta &= -i\nu + O_1(\chi,|\zeta|^2), \end{aligned}$$

such that solutions of the transformed equations are given by

$$\begin{aligned} \xi(t) &= \xi^0 e^{t\alpha}, & \eta(t) &= \eta^0 e^{-t\alpha},\\ \zeta(t) &= \zeta^0 e^{t\beta}, & \bar{\zeta} &= \bar{\zeta}^0 e^{-t\beta}, \end{aligned}$$

where ξ^0,η^0 and ζ^0 are determined from the initial conditions and $\bar{\zeta}$ is the complex conjugate of ζ. Furthermore, the coefficients of α and β are real and complex, respectively, from which it follows that the functions $\xi\eta = x_1y_1 + O_3(x,y)$ and $|\zeta|^2 = x_2^2+y_2^2+O_3(x,y)$ are local integrals, as are α and β.

Finally, the transformation of the Hamiltonian function has the form

$$K(\xi,\eta,\zeta,\bar{\zeta}) = H(x,y) = \lambda\xi\eta + \tfrac{1}{2}|\zeta|^2 + O_2(\chi,|\zeta|^2),$$

and in particular depends only on the variables $\chi=\xi\eta$ and $|\zeta|^2$.

2.7 Geometry of Solutions near the Equilibria

Now we analyze the linearized equations (2.6.4). It is straightforward to find that the eigenvalues of this linear system have the form $\pm\lambda$ and $\pm i\nu$, where λ and ν are positive constants. The corresponding eigenvectors are

$$\begin{aligned} u_1 &= (1,-\sigma,\lambda,-\lambda\sigma),\\ u_2 &= (1,\sigma,-\lambda,-\lambda\sigma),\\ w_1 &= (1,-i\tau,i\nu,\nu\tau),\\ w_2 &= (1,i\tau,-i\nu,\nu\tau), \end{aligned}$$

where σ and τ are constants with $\sigma>0$ and $\tau<0$.

Eigenvalues. It can be shown that the characteristic polynomial for the linearized equations (2.6.4) written in matrix form

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{v}_x \\ \dot{v}_y \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a & 0 & 0 & 2 \\ 0 & -b & -2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ v_x \\ v_y \end{pmatrix} = A \begin{pmatrix} x \\ y \\ v_x \\ v_y \end{pmatrix},$$

is given by

$$p(\beta) = \beta^4 + (2 - \bar{\mu})\beta^2 + (1 + \bar{\mu} - 2\bar{\mu}^2).$$

Let $\alpha = \beta^2$, then the roots of $p(\alpha) = 0$ are as follows

$$\alpha_1 = \frac{\bar{\mu} - 2 + \sqrt{9\bar{\mu}^2 - 8\bar{\mu}}}{2}, \qquad \alpha_2 = \frac{\bar{\mu} - 2 - \sqrt{9\bar{\mu}^2 - 8\bar{\mu}}}{2}.$$

Since the last term of $p(\alpha) = 0$ is equal to $-ab$ which is negative, this quadratic equation must have one positive and one negative root. So, we have $\alpha_1 > 0$ and $\alpha_2 < 0$. Therefore, the eigenvalues of the linearized equations are of the form $\pm\lambda$ and $\pm i\nu$, where $\lambda = \sqrt{\alpha_1}$ and $\nu = \sqrt{-\alpha_2}$.

Eigenvectors. Let $v = (k_1, k_2, k_3, k_4)$ be an eigenvector of the linearized equations. If β is an eigenvalue, then $Av = \beta v$ and we have the following relations

$$\begin{aligned} k_3 &= \beta k_1, & ak_1 + 2k_4 &= \beta k_3, \\ k_4 &= \beta k_2, & -bk_2 - 2k_3 &= \beta k_4. \end{aligned}$$

Notice that $k_1 \neq 0$, otherwise $k_2 = k_3 = k_4 = 0$ and $v = 0$. Thus, k_1 may be taken to be 1 and the equations relating the components of v indicate that v may have the form

$$v = (1, k_2, \beta, \beta k_2),$$

and that

$$\begin{aligned} a + 2\beta k_2 &= \beta^2, \\ -bk_2 - 2\beta &= \beta^2 k_2. \end{aligned}$$

First let $\beta = \lambda$ and then $\beta = -\lambda$ to obtain

$$\begin{aligned} u_1 &= (1, k_2, \lambda, \lambda k_2), \\ u_2 &= (1, k_2', -\lambda, -\lambda k_2'), \end{aligned}$$

where

$$\begin{aligned} a + 2\lambda k_2 &= \lambda^2, \\ -bk_2 - 2\lambda &= \lambda^2 k_2, \\ a - 2\lambda k_2' &= \lambda^2, \\ -bk_2' + 2\lambda &= \lambda^2 k_2'. \end{aligned}$$

The first and the third equations show that $k_2 = -k_2'$, and, denoting $k_2' = \sigma$, the second and fourth give

$$\sigma = \frac{2\lambda}{\lambda^2 + b} > 0. \tag{2.7.1}$$

Similarly, taking $\beta = i\nu$, then $\beta = -i\nu$, we obtain

$$\begin{aligned} w_1 &= (1, -i\tau, i\nu, \nu\tau), \\ w_2 &= (1, i\tau, -i\nu, \nu\tau), \end{aligned}$$

where

$$\tau = -\left(\frac{\nu^2 + a}{2\nu}\right) < 0. \tag{2.7.2}$$

Eigenvectors as Axes for New Coordinate System. To better understand the orbit structure on the phase space, we make a linear change of coordinates with the eigenvectors, u_1, u_2, w_1, w_2, as the axes of the new system. Using the corresponding new coordinates $(\xi, \eta, \zeta_1, \zeta_2)$, the differential equations assume the simple form

$$\begin{aligned} \dot{\xi} &= \lambda\xi, \\ \dot{\eta} &= -\lambda\eta, \\ \dot{\zeta}_1 &= \nu\zeta_2, \\ \dot{\zeta}_2 &= -\nu\zeta_1, \end{aligned} \tag{2.7.3}$$

and the energy function (2.6.5) becomes

$$E_l = \lambda\xi\eta + \frac{\nu}{2}(\zeta_1^2 + \zeta_2^2). \tag{2.7.4}$$

Solutions of the equations (2.7.3) can be conveniently written as

$$\begin{aligned} \xi(t) &= \xi^0 e^{\lambda t}, \\ \eta(t) &= \eta^0 e^{-\lambda t}, \\ \zeta(t) &= \zeta_1(t) + i\zeta_2(t) = \zeta^0 e^{-i\nu t}, \end{aligned} \tag{2.7.5}$$

where the constants ξ^0, η^0 and $\zeta^0 = \zeta_1^0 + i\zeta_2^0$ are the initial conditions. These linearized equations admit integrals in addition to the energy function (2.7.4); namely, the functions $\eta\xi$ and $|\zeta|^2 = \zeta_1^2 + \zeta_2^2$ are both constant along solutions.

Phase Space of the Equilibrium Region. For positive ε and c, the region $\mathcal{R}$, which is determined by

$$E_l = \varepsilon, \quad \text{and} \quad |\eta - \xi| \le c, \tag{2.7.6}$$

which is homeomorphic to the product of a two-sphere and an interval; namely, for each fixed value of $\eta-\xi$ on the interval $I=[-c,c]$, the equation $E_l=\varepsilon$ determines the two-sphere

$$\frac{\lambda}{4}(\eta+\xi)^2+\frac{\nu}{2}(\zeta_1^2+\zeta_2^2)=\varepsilon+\frac{\lambda}{4}(\eta-\xi)^2.$$

The bounding sphere of $\mathcal{R}$ for which $\eta-\xi=-c$ will be called n_1, and that where $\eta-\xi=c$, n_2 (see Figure 2.7.1). We shall call the set of points on each bounding sphere where $\eta+\xi=0$ the *equator*, and the sets where $\eta+\xi>0$ or $\eta+\xi<0$ will be called the *north and south hemispheres*, respectively.

Flow in the Equilibrium Region. To analyze the flow in $\mathcal{R}$ one simply considers the projections on the η-ξ plane and ζ planes, respectively. In the first case we see the standard picture of an unstable critical point, and in the second, of a center. Figure 2.7.1 schematically illustrates the flow in the η-ξ plane. The coordinate axes have been tilted by 45° in order to correspond to the direction of the flow in later figures. In Figure 2.7.1(a), $\mathcal{R}$ itself projects to a set bounded on two sides by the hyperbola $\eta\xi=\varepsilon/\lambda$, the thick solid hyperbolic segments on the top and bottom, (corresponding to $|\zeta|^2=0$, see (2.7.4)). $\mathcal{R}$ is bounded on two other sides by the line segments $\eta-\xi=\pm c$, the dotted vertical lines at left and right in Figure 2.7.1(a), which correspond to the bounding spheres, n_1 and n_2, respectively.

Since $\eta\xi$ is an integral of the equations in $\mathcal{R}$, the projections of orbits in the η-ξ plane move on the branches of the corresponding hyperbolas $\eta\xi=$ constant, except in the case $\eta\xi=0$ (where $\eta=0$ or $\xi=0$). If $\eta\xi>0$, the branches connect the bounding line segments $\eta-\xi=\pm c$ and if $\eta\xi<0$, they have both end points on the same segment. A check of equation (2.7.5) shows that the orbits move as indicated by the arrows in Figure 2.7.1(b).

To interpret Figure 2.7.1(b) as a flow in $\mathcal{R}$, notice that each point in the projection corresponds to a circle in $\mathcal{R}$ given by the "radius" variable $\rho=|\zeta|^2=$ constant. Recall from (2.7.4) that $|\zeta|^2=\frac{2}{\nu}(\varepsilon-\lambda\eta\xi)$. Of course, for points on the bounding hyperbolic segments ($\eta\xi=\varepsilon/\lambda$), the constant is zero so that the circle collapses to a point. Thus, the segments of the lines $\eta-\xi=\pm c$ in the η-ξ projection correspond to the two-spheres bounding $\mathcal{R}$. This is because each corresponds to a circle crossed with an interval where the two end circles are pinched to a point.

We distinguish nine classes of orbits grouped into the following four categories:

1. The point at the origin in Figure 2.7.1(b), $\xi=\eta=0$, corresponds to a ***periodic*** orbit in $\mathcal{R}$, known as the ***Lyapunov orbit***[5] (in, e.g., Szebehely [1967]).

[5]Aleksandr Mikhailovich Lyapunov (1857–1918) was a Russian mathematician, mechanician and physicist who worked on differential equations, celestial mechanics, and

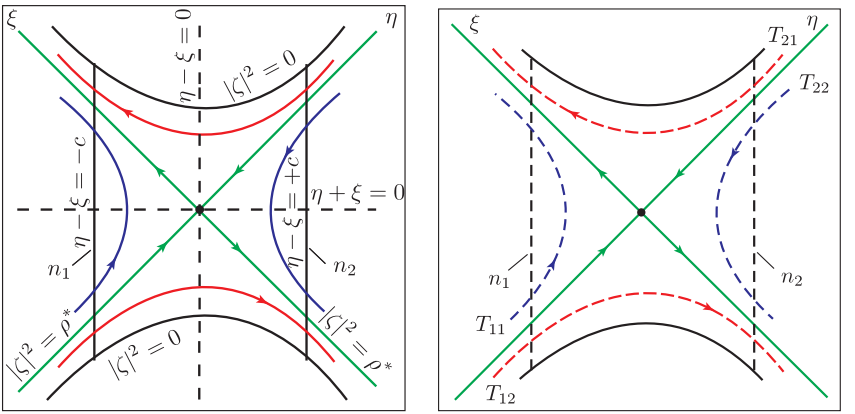


FIGURE 2.7.1. **The projection onto the η-ξ plane** of orbits near the equilibrium point (note, axes tilted 45°). (a) The equilibrium region, $\mathcal{R}$, is bounded by the thick hyperbolic segments at top and bottom and the dotted vertical segments at left and right. At the origin is the periodic orbit in $\mathcal{R}$. The thick lines with arrows pointing toward or away from the origin are trajectories asympototically winding onto the periodic orbit. See the text for further descriptions. (b) Four additional trajectories are shown. The labeling T_{ij} denotes the path of a particle which entered $\mathcal{R}$ through n_i and exited through n_j. Two transit orbits, T_{12} and T_{21}, and two non-transit orbits, T_{11} and T_{22}, are shown.

2. The four half-open segments on the axes, $\eta\xi = 0$ (or equivalently $|\zeta|^2 = \rho^*$ where $\rho^* = 2\varepsilon/\nu$), correspond to four cylinders of orbits asymptotic to this periodic solution either as time increases ($\xi = 0$) or as time decreases ($\eta = 0$). These are called ***asymptotic*** orbits. The are drawn as the thick lines with arrows pointing toward or away from the origin in Figures 2.7.1(a) and (b).

3. The hyperbolic segments determined by $\eta\xi = \text{constant} > 0$ (or equivalently $|\zeta|^2 < \rho^*$) correspond to two cylinders which cross $\mathcal{R}$ from one bounding sphere to the other, meeting both in the same hemisphere; the north one if they go from $\eta - \xi = +c$ to $\eta - \xi = -c$, the south one in the other case. Since these orbits transit from one region to another, we call them ***transit*** orbits. The two trajectories labeled T_{12} and T_{21} in Figure 2.7.1(b) are transit orbits.

4. Finally the hyperbolic segments determined by $\eta\xi = \text{constant} < 0$ ($|\zeta|^2 > \rho^*$) correspond to two cylinders of orbits in $\mathcal{R}$ each of

probability theory. His main preoccupations were the stability of equilibrium and the motion of mechanical systems, the model theory for the stability of uniform turbulent liquids, and the motion of particles under the influence of gravity.

which runs from one hemisphere to the other hemisphere on the same bounding sphere. Thus if $\xi > 0$, the sphere is n_1 ($\eta - \xi = -c$) and orbits run from the south ($\eta + \xi < 0$) to the north ($\eta + \xi > 0$) hemisphere while the converse holds if $\xi < 0$, where the sphere is n_2. Since these orbits return to the same region, we call them ***non-transit*** orbits. See the two trajectories labeled T_{11} and T_{22} in Figure 2.7.1(b).

McGehee Representation of the Equilibrium Region. McGehee [1969], building on the work of Conley [1968], proposed a representation which makes it easier to visualize the region $\mathcal{R}$. Recall that $\mathcal{R}$ is homeomorphic to $S^2 \times I$. In McGehee [1969], it is represented by a spherical annulus, as shown in Figure 2.7.2(b).

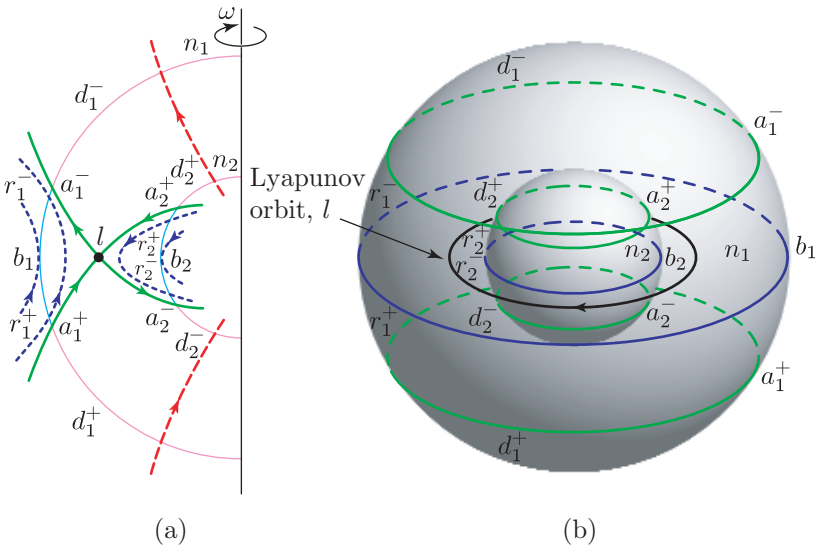


FIGURE 2.7.2. **McGehee representation of the equilibrium region.** (a) The cross section of the flow in the $\mathcal{R}$ region of the energy surface. (b) The McGehee representation of the flow in the region $\mathcal{R}$. See the text for details.

Figure 2.7.2(a) is a cross section of $\mathcal{R}$. Notice that this cross section is qualitatively the same as the illustration in Figure 2.7.1(b). The full picture (Figure 2.7.2(b)) is obtained by rotating this cross section about the indicated axis ω. The following classifications of orbits correspond to the previous four categories:

1. There is an unstable *periodic* orbit l, the Lyapunov orbit in the region $\mathcal{R}$.

2. Again let n_1, n_2 be the bounding spheres of region $\mathcal{R}$, and let n denote either n_1 or n_2. We can divide n into two hemispheres: n^+, where

the flow enters $\mathcal{R}$, and n^-, where the flow leaves $\mathcal{R}$. We let a^+ and a^- (where $|\zeta|^2 = \rho^*$) be the intersections with n of the cylinders of orbits *asymptotic* to the unstable periodic orbit l. As shown in Figure 2.7.2(b), a^+ appears as a circle in n^+, and a^- appears as a circle in n^-.

3. If we let d^+ be the spherical cap (where $|\zeta|^2 < \rho^*$) in n^+ bounded by a^+, shown in Figure 2.7.2(b), then the *transit* orbits entering $\mathcal{R}$ on d^+ exit on d^- of the other bounding sphere. Similarly, letting d^- ($|\zeta|^2 < \rho^*$) be the spherical cap in n^- bounded by a^-, the transit orbits leaving on d^- have come from d^+ on the other bounding sphere.

4. Note that the intersection b of n^+ and n^- is a circle of tangency points. Orbits tangent at this circle "bounce off," i.e., do not enter $\mathcal{R}$ locally. Moreover, if we let r^+ be a spherical zone which is bounded by a^+ and b, then *non-transit* orbits entering $\mathcal{R}$ on r^+ (where $|\zeta|^2 > \rho^*$) exit on the same bounding sphere through r^- (where $|\zeta|^2 > \rho^*$) which is bounded by a^- and b.

Invariant Manifold Tubes as Separatrices. The key observation here is that the asymptotic orbits are pieces of the stable and unstable cylindrical manifolds of the Lyapunov orbit and they separate two distinct types of motion: transit orbits and non-transit orbits. The transit orbits, passing from one realm to another, are those inside the cylindrical manifold, or ***tube***. The non-transit orbits, which bounce back to their realm of origin, are those outside the tube. This observation will be important for the numerical construction of interesting orbits in Chapters 4 and 5.

2.8 Flow Mappings in the Equilibrium Region

We now observe that on the two bounding spheres, each of the hemispheres $n^\pm$ is transverse to the flow. It follows that the flow in $\mathcal{R}$ defines four mappings—two between pairs of spherical caps $d^\pm$ and two between pairs of spherical zones $r^\pm$ (as in Llibre, Martinez, and Simó [1985]):

$$\psi_1 : d_1^+ \to d_2^-, \qquad \psi_2 : d_2^+ \to d_1^-, \tag{2.8.1}$$

$$\psi_3 : r_1^+ \to r_1^-, \qquad \psi_4 : r_2^+ \to r_2^-. \tag{2.8.2}$$

The four mappings are diffeomorphisms. Furthermore, all these mappings preserve the radius variable $\rho = |\zeta|^2$ since this is an integral in $\mathcal{R}$.

The Infinite Twisting of the Mappings. After computing from the solution (2.7.5) that

$$\frac{d}{dt} \arg \zeta = -\nu, \tag{2.8.3}$$

we see that the change in the argument of ζ for each of these mappings ψ_i is approximately proportional to the negative of the time required to go from domain to range.

Also, this time approaches infinity as the flow approaches the circle a^+ ($|\zeta|^2 \to \rho^*$), since on the circle a^+ (where $|\zeta|^2 = \rho^*$) the orbits are asymptotic to the unstable periodic solution l. The proof is quite straightforward. Take ψ_2 as an exmaple. According to equations (2.7.5), we have $\xi(0) = \xi^0, \eta(0) = \eta^0$ on d_2^+ where η^0 and ξ^0 are both positive and

$$\eta^0 - \xi^0 = +c.$$

Similarly, if T is the time required to go from domain to range, then $\xi(T) = \xi^0 e^{\lambda T}$ and $\eta(T) = \eta^0 e^{-\lambda T}$ on d_1^-, where

$$\eta(T) - \xi(T) = \eta^0 e^{-\lambda T} - \xi^0 e^{\lambda T} = -c.$$

Eliminating c from the two above equations and solving for T, we obtain

$$T = \frac{1}{\lambda} \ln \frac{\eta^0}{\xi^0}.$$

Moreover, the energy integral (2.7.4) gives

$$\xi^0 \eta^0 = \frac{\varepsilon}{\lambda} - \frac{\nu}{2\lambda}|\zeta|^2 = \frac{\nu}{2\lambda}\left(\frac{2\varepsilon}{\nu} - |\zeta|^2\right) = \frac{\nu}{2\lambda}(\rho^* - \rho).$$

Hence,

$$T = \frac{1}{\lambda}\left(\ln \frac{2\lambda(\eta^0)^2}{\nu(\rho^* - \rho)}\right) = \frac{1}{\lambda}\left(\ln \frac{2\lambda(\eta^0)^2}{\nu} - \ln(\rho^* - \rho)\right)$$

where the last term determines the order of the required transit time.

These facts imply that arbitrary circles with radius variable $\rho = |\zeta|^2$ in the domain of the mappings are rotated by an amount that decreases to minus infinity as $\rho \to \rho^*$. Hence, the behavior of the flow in $\mathcal{R}$ should be obtained by adding some spiraling to the arrows given in Figure 2.7.2(a).

In Chapters 3 and 5, we will need a simple geometric consequence of the above observation on spiraling stated in terms of "abutting arcs" in the domain, or range of ψ_i. Namely, an arc lying in the closure of one of these sets ($d^\pm$ and $r^\pm$) is called an abutting arc if it is in the set itself except for one end point in the circle $a^\pm$. See Figure 2.8.1. For example, let γ_1 be an abutting arc in the domain d_1^+ of ψ_1 with one end point P_1 in a_1^+. Let δ_1 be another abutting arc in the range d_2^- of ψ_1 such that one of its end point Q_1 is in a_2^-. Then $\psi_1(\gamma_1)$ is an arc spiraling towards a_2^- and cutting δ_1 an infinite number of times in any neighborhood of the point of abutment Q_1.

This follows directly from the infinite twisting of the mappings ψ_1; namely, the image of γ_1 spirals infinitely many times around and down to a_2^- in the range.

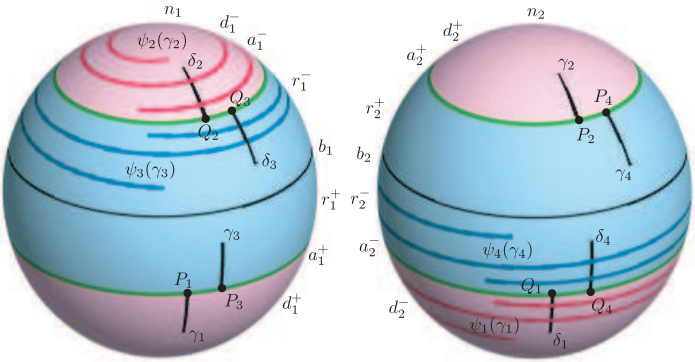


FIGURE 2.8.1. Spiraling of the images of arcs γ_i.

Similarly, let γ_i be an abutting arc in the domain of ψ_i with one end point P_i in a_2^+, a_1^+, a_2^+ for $i = 2, 3, 4$, respectively. Let δ_i be another abutting arc in the range of ψ_i such that one of its end points Q_i is in a_1^-, a_1^-, a_2^- respectively. Then $\psi_i(\gamma_i)$ is an arc spiraling towards a_1^-, a_1^-, a_2^-, respectively and cutting δ_i an infinite number of times in any neighborhood of the point of abutment Q_i.

2.9 Trajectories in the Neck Region

Having studied the orbit structure in the equilibrium region $\mathcal{R}$ and its projection on the η-ξ plane, we now examine briefly the appearance of orbits in position space, that is, in the x-y plane. In position space, R appears as the neck region connecting two realms, so trajectories in $\mathcal{R}$ will be projected to tarjectories in the neck region.

Recall from §2.6 that the ξ and η coordinate axes are the eigenvectors $u_1 = (1, -\sigma, \lambda, -\lambda\sigma)$ and $u_2 = (1, \sigma, -\lambda, -\lambda\sigma)$, respectively. Their projection on the x-y plane, $\bar{u}_1 = (1, -\sigma)$ and $\bar{u}_2 = (1, \sigma)$, plays an important role in the study of the appearance of orbits on the position space.

The image of a tilted projection of $\mathcal{R}$ on the x-y plane provides the right mental picture. To build physical intuition regarding the flow in the equilibrium region, it is important to study the projection of the different classes of orbits on the x-y plane. Here, we summarize the main results of Conley [1968].

Recall from §2.6 that the eigenvalues of the linear system (2.6.4) are $\pm\lambda$ and $\pm i\nu$ with corresponding eigenvectors u_1, u_2, w_1, w_2. Thus, the general (real) solution has the form

$$\mathbf{z}(t) = (x(t), y(t), \dot{x}(t), \dot{y}(t)) = \alpha_1 e^{\lambda t} u_1 + \alpha_2 e^{-\lambda t} u_2 + 2\mathrm{Re}(\beta e^{i\nu t} w_1), \quad (2.9.1)$$

where α_1, α_2 are real and $\beta = \beta_1 + i\beta_2$ is complex. Notice that (2.9.1), while slightly more complicated, is essentially the same as (2.7.5).

Upon inspecting this general solution, we see that the solutions on the energy surface fall into different classes depending upon the limiting behavior of $x(t)$ (the x coordinate of $\mathbf{z}(t)$) as t tends to plus or minus infinity. Notice that

$$x(t) = \alpha_1 e^{\lambda t} + \alpha_2 e^{-\lambda t} + 2(\beta_1 \cos \nu t - \beta_2 \sin \nu t). \quad (2.9.2)$$

Thus, if $t \to +\infty$, then $x(t)$ is dominated by its α_1 term. Hence, $x(t)$ tends to minus infinity (staying on the left-hand side), is bounded (staying around the equilibrium point), or tends to plus infinity (staying on the right-hand side) according to $\alpha_1 < 0, \alpha_1 = 0, \alpha_1 > 0$. See Figure 2.9.1. The same statement holds if $t \to -\infty$ and α_2 replaces α_1. Different combinations of the signs of α_1 and α_2 will give us again the same nine classes of orbits which can be grouped into the same four categories:

1. If $\alpha_1 = \alpha_2 = 0$, we obtain a *periodic* solution which is a Lyapunov orbit. It has been proven in Conley [1968] that this periodic orbit, shown in Figure 2.9.1, projects onto the x-y plane as an ellipse with major axis of length $2|\tau|\sqrt{\varepsilon/\kappa}$ in the direction of the y-axis, and minor axis of length $2\sqrt{\varepsilon/\kappa}$ in the direction of the x-axis. The orientation of the orbit is clockwise. Here κ $(= -a + b\tau^2 + \nu^2 + \nu^2\tau^2)$ is a positive constant. Note that the size of the ellipse goes to zero with ε.

2. Orbits with $\alpha_1\alpha_2 = 0$ are *asymptotic* orbits. They are asymptotic to the periodic Lyapunov orbit. It has been proven in Conley [1968] that the asymptotic orbits with $\alpha_1 = 0$ project into the strip S_1 in the x-y plane centering around $\bar{u}_2$ and bounded by the lines

$$y = \sigma x \pm 2\sqrt{\varepsilon(\sigma^2 + \tau^2)/\kappa}. \quad (2.9.3)$$

   Similarly, asymptotic orbits with $\alpha_2 = 0$ project into the strip S_2 centering around $\bar{u}_1$ and bounded by the lines

$$y = -\sigma x \pm 2\sqrt{\varepsilon(\sigma^2 + \tau^2)/\kappa}. \quad (2.9.4)$$

   Notice that the width of the strips goes to zero with ε.

3. Orbits with $\alpha_1\alpha_2 < 0$ are *transit* orbits because they cross the equilibrium region R from $-\infty$ (the left-hand side) to $+\infty$ (the right-hand side) or vice versa.

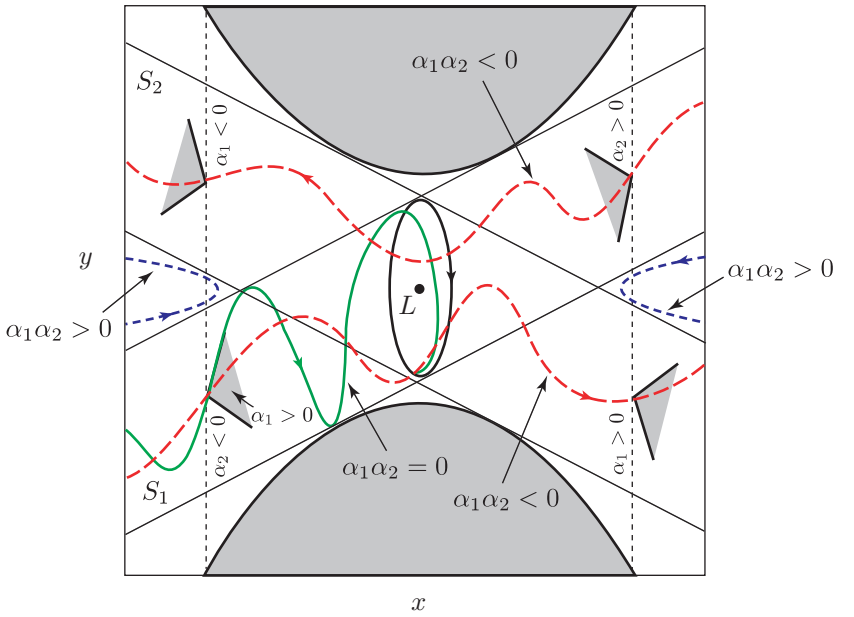


FIGURE 2.9.1. **The flow in the equilibrium region R of position space.** Shown are the periodic orbit (ellipse); a typical asymptotic orbit winding onto the periodic orbit; two transit orbits (dashed); and two non-transit orbits (dotted). See the text for an explanation of the labeling.

4. Orbits with $\alpha_1\alpha_2 > 0$ are *non-transit* orbits.

To study the projection of these last two categories of orbits, Conley [1968] proved a couple of propositions which allows one to determine at each position (x, y) the "wedge" of velocities (if any) in which $\alpha_1\alpha_2 < 0$. See the shaded wedges in Figure 2.9.1. Since a detailed study will draw us too far afield, we simply state some of the main observations.

In Figure 2.9.1, S_1 and S_2 are the two strips mentioned above. Outside of each strip $S_i, i = 1, 2$, the sign of α_i is independent of the direction of the velocity. These signs can be determined in each of the components of the equilibrium region R complementary to both strips. For example, in the left-most central components, both α's are negative, while in the right-most central components both α's are positive. Therefore, $\alpha_1\alpha_2 > 0$ in both components and only non-transit orbits project onto these two components.

Inside the strips the situation is more complicated since in $S_i, i = 1, 2$, the signs of α_i depends on the direction of the velocity. For simplicity we have indicated this dependence only on the two vertical bounding line segments in Figure 2.9.1. For example, consider the intersection of strip S_1 with the left-most vertical line. On the subsegment so obtained there is at each point

a wedge of velocity in which α_1 is positive. The sign of α_2 is always negative on this subsegment, so that orbits with velocity interior to the wedge are transit orbits ($\alpha_1\alpha_2 < 0$). Of course, orbits with velocity on the boundary of the wedge are asymptotic ($\alpha_1\alpha_2 = 0$), while orbits with velocity outside of the wedge are non-transit. Here, only a transit and asymptotic orbit are illustrated. The situation on the remaining three subsegments is similar.

The Flow in the Equilibrium Region. In summary, the phase space in the equilibrium region can be partitioned into four categories of distinctly different kinds of motion (see Figures 2.7.1, 2.7.2 and 2.9.1): the periodic Lyapunov orbits, asymptotic orbits, transit orbits, and, finally, non-transit orbits.

3

Heteroclinic Connection and Global Orbit Structure

3.1 Introduction.

As mentioned in Chapter 2, near one of the equilibrium points L_1 or L_2, there is a family of unstable periodic orbits. For energy values in case 3, the energy surface contains exactly one of these periodic solutions near each libration point. As dynamical systems theory suggests (see, for example, Wiggins [2003]), to understand fully the global dynamics of the flow, one should examine structures like homoclinic and heteroclinic orbits connecting these L_1 and L_2 Lyapunov orbits to themselves.

In §2.7, the local orbit structure near the libration points was shown to give (i) periodic orbits (the Lyapunov orbits), (ii) pieces of the stable and unstable manifolds of these periodic orbits, (iii) transit and (iv) non-transit orbits. In this chapter, we explore how these local structures are connected globally. Our goal is to show how homoclinic orbits in the interior realm are connected to the homoclinic orbits in the exterior realm by a heteroclinic cycle in the Jupiter realm. We refer to the union of these three structures as a ***chain***. An example is given in Figure 3.1.1.

The story is completed later in the chapter when this dynamical chain structure is used to show the existence of complex and interesting trajectories, some of which have been observed in actual comet trajectories, as discussed in Koon, Lo, Marsden, and Ross [2000].

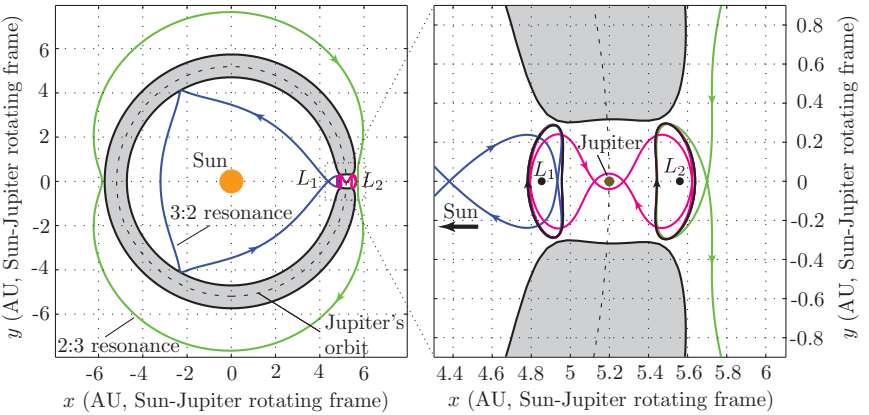


FIGURE 3.1.1. **A homoclinic-heteroclinic chain** corresponding to the Jupiter comet Oterma. The p.o.'s about L_1 and L_2 are black. Their homoclinic orbits are labeled with the mean motion resonances with which they are associated. The orbit homoclinic to the L_1 p.o. in the interior realm is labeled as the "3:2 resonance," since the particle (or comet) goes around the Sun 3 times while Jupiter goes around the sun 2 times as seen in an inertial frame. The orbit homoclinic to the L_2 p.o. in the exterior realm is similarly labeled as the "3:2 resonance." The pair of heteroclinic orbits connecting the L_1 and L_2 p.o.'s is also shown. These four structures together form a *chain*. We say this chain corresponds to the comet Oterma because this comet follows a trajectory close to this chain, as will be elaborated upon in this chapter. Distances are given in Astronomical Units (AU, about 150 million km).

Trajectories can be categorized by their *itinerary* with respect to the three realms, as shown in Figure 3.1.2. In other words, an itinerary such as $(\ldots, S; J, X, J, \ldots)$ lists the future and past whereabouts of a spacecraft or comet; currently in the J (or Jupiter) realm, it is heading to the X (or exterior) realm and came from the S (or Sun) realm.

What we find is an invariant set of orbits, to each of which we can attach an itinerary (e.g., $(\ldots, X; J, S, J, \ldots)$ in the informal notation) describing the future and past history of the orbit for all time. At the end of this chapter, we prove that any permissible itinerary which can be written corresponds to an actual solution of the three-body problem.

Guided by ideas laid down in this chapter, we numerically construct some interesting trajectories in Chapters 4 and 5.

A Note of Caution to the Reader. The mathematics in this chapter can be challenging and is not necessary for the rest of the book. The reader interested mostly in mission design can skip this chapter if desired without breaking the flow of the book.

Organization of the Chapter. In more detail, this chapter discusses the following topics. In §3.2, we discuss some of the results from Conley [1968] and McGehee [1969], which prove the existence of ***homoclinic*** orbits

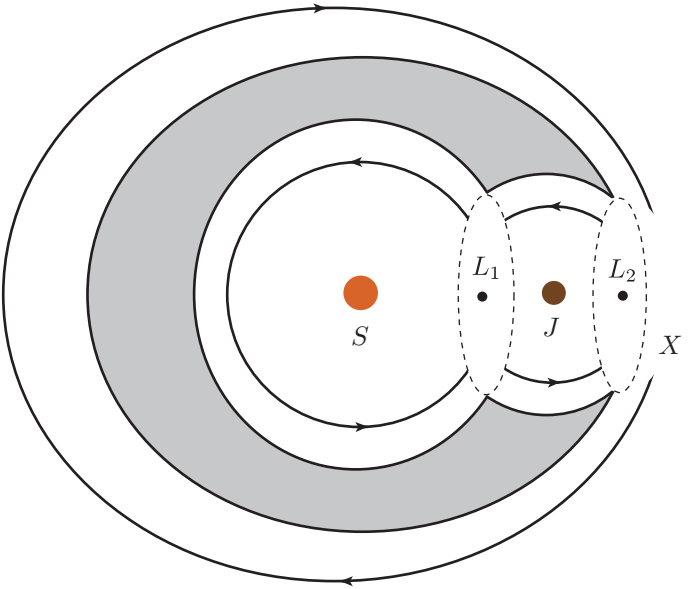


FIGURE 3.1.2. Trajectories can be categorized by their itinerary with respect to the three realms. For example, for a comet in the Sun-Jupiter system, the three realms are S, J, and X. We prove a theorem given at the end of the chapter which states that any an permissible itinerary (e.g., $(\ldots, X; J, S, J, \ldots)$ in the informal notation) corresponds to a solution of the three-body problem. Guided by the proof, we also outline a procedure for finding such trajectories, which will be useful in later chapters.

in both the interior and exterior realms. These are the orbits which are both forward and backward asymptotic to an unstable Lyapunov orbit. The heart of the proof is the construction of a function which counts the number of times an orbit segment with endpoints near the Lyapunov orbit winds around a solid torus.

We discuss in §3.3 the main results in Llibre, Martinez, and Simó [1985] on the transversality of the invariant manifolds for the L_1 Lyapunov orbit. In dynamical systems theory, the property of being doubly asymptotic to a periodic orbit is described (and more quantitatively handled) by saying that the orbit is in both the *stable* and *unstable* manifold of the periodic orbit, or that the homoclinic orbit is in the intersection of the stable and unstable manifolds of the periodic orbit. One of the most important issues which arises in this context is the transversality of the intersection. The presence of transversality will allow us to draw many profound conclusions about the orbit structure of the system under study.

Since neither Conley [1968] nor McGehee [1969] was able to settle this issue, Llibre, Martinez, and Simó [1985] spent their major effort in proving analytically that the intersection is indeed transversal under appropriate conditions, at least in the interior realm. We summarize their results.

The theorems given in §3.3 are cited only for guidance on how to construct the transversal homoclinic orbits numerically. In §3.4 we compute intersections of stable and unstable manifolds of L_1 and L_2 p.o.'s to numerically demonstrate the existence of transversal homoclinic orbits in both the interior and exterior realms.

In §3.5 we use similar computational methods to numerically demonstrate the existence of transversal ***heteroclinic*** orbits in the Jupiter realm which connect asymptotically the L_1 and L_2 Lyapunov orbits. A heteroclinic orbit, also known as a ***heteroclinic connection***, is an orbit lying in the intersection of the stable manifold of one periodic orbit and the unstable manifold of another periodic orbit. As discussed in Chapter 2, since the PCR3BP is a Hamiltonian system with two degrees of freedom, its energy manifold is three-dimensional. From the work of Conley [1968], it was known that both the stable and unstable manifolds of the p.o.'s around L_1 and L_2 are two-dimensional. Hence, a dimension count suggests, but does not prove, the existence of such a heteroclinic connection. Careful numerical investigations allow us to show this connection is indeed present, as well as to isolate and study it.

It is worth noting that, inspired by these numerical demonstrations which were first reported in Koon, Lo, Marsden, and Ross [2000], two teams of authors have rigorously proven the existence of transversal homoclinic and heteroclinic orbits (Wilczak and Zgliczyński [2003] and Kirchgraber and Stoffer [2004]). These are "computer-assisted proofs" using interval analysis methods, and they further verify the claims we make in this chapter.

In §3.6, we numerically demonstrate that, within an appropriate range of energy values, there exist chains of two homoclinic orbits and a symmetric heteroclinic cycle, as in Figure 3.1.1. The existence of these chains will be used in §3.7 to construct a suitable Poincaré map which will allow us to classify as well as organize distinctively different types of global motions of the PCR3BP in terms of ultimate behavior with respect to the equilibrium points.

In §3.8 and §3.9, we extend the symbolic dynamics results of Llibre, Martinez, and Simó [1985] to our situation and construct a set of bi-infinite sequences with two families of symbols.

In §3.10, we state the main theorem of this chapter and discuss its implications. The theorem gives the global orbit structure of the PCR3BP in a neighborhood of a chain of homoclinic orbits and a symmetric heteroclinic cycle.

3.2 Existence of Orbits Homoclinic to the Lyapunov Orbit

In order to facilitate explanation of concepts in this chapter, we urge the reader to keep a concrete physical example in mind—the motion of a comet in the gravity field of the Sun and Jupiter.

Energy Manifold and Hill's Region. Recall from §2.4 that the motion of a comet with energy e is restricted to a three-dimensional energy surface $\mathcal{M}(\mu, e)$. The projection of the energy surface onto position space, $M(\mu, e)$, is the region of possible motion known as the Hill's region. The first four cases of possible motion, depending on energy, are given in Figure 2.4.2. For case 3, the region of possible motion contains necks around both L_1 and L_2 and the comet can transit from the interior realm to the exterior realm and vice versa. This is the case of most interest to us.

Orbit Segments Winding around a Solid Torus. From McGehee [1969], we know that the energy surface is broken up further into regions bounded by invariant tori. These invariant tori $\mathcal{A}_1$ and $\mathcal{A}_2$ project onto the darkly shaded annuli A_1 and A_2, respectively, shown for case 3 in Figure 3.2.1(a).

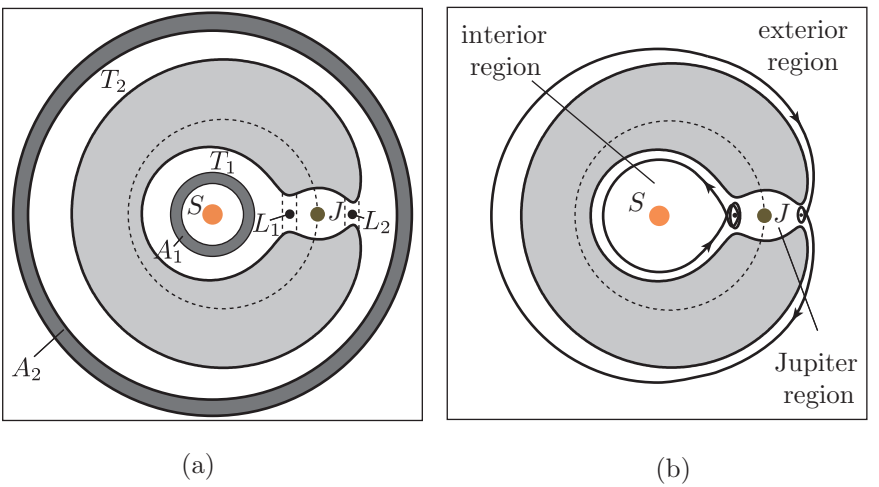


FIGURE 3.2.1. (a) The projection of invariant tori (darkly shaded) on position space for case 3. (b) Homoclinic orbits in the interior and exterior realms.

These annuli separate the Hill's region into sets corresponding to the invariant sets in the energy surface. It is interesting to note that for all of the cases, the singularities corresponding to the center of the Sun and Jupiter are separated from each other by an invariant torus (although we show

only case 3), thus making it impossible for the comet to pass from a region arbitrarily close to the Sun to a region arbitrarily close to Jupiter. Similarly, Jupiter is separated from infinity by an invariant torus. We consider the regions of the energy surface projecting to the area between the two darkly shaded annuli, A_1 and A_2, i.e., the region containing Jupiter. The theorems of McGehee given below show that all orbits leaving the vicinity of one of the unstable periodic orbits proceed around one of the annuli T_1 or T_2, projections of *solid tori* $\mathcal{T}_1$ or $\mathcal{T}_2$, before returning to that vicinity. The direction of motion is the same for all orbits, counterclockwise in the interior realm and clockwise in the exterior realm.

In Chapter 2, we studied the regions near the unstable periodic orbits to obtain a qualitative picture of the asymptotic orbits. Here we combine this picture of asymptotic orbits with the fact that orbits in T_1 or T_2 wind around in one direction *to construct homoclinic orbits* in both the interior and exterior realms, shown schematically in Figure 3.2.1(b).

We note that the following theorems do not literally apply to the system parameters (μ) and energies (e) of interest to us. But they are a useful guide and numerical experiments reveal that the qualitative results they suggest hold over a larger set of system parameters and energies than is proven.

The Theorems of McGehee. To precisely state the theorems, we must first divide up the Hill's region and the energy surface. From §2.5, for small μ the two equilibrium points occur at an approximate distance $r_h = (\mu/3)^{\frac{1}{3}}$ on either side of Jupiter. We isolate these points by drawing vertical lines on each side of them, i.e., lines at $(1-\mu \pm c_1 r_h, 0)$ and $(1-\mu \pm b_1 r_h, 0)$, where $b_1 < 1 < c_1$. This divides the Hill's region into five sets as shown in Figure 3.2.2.

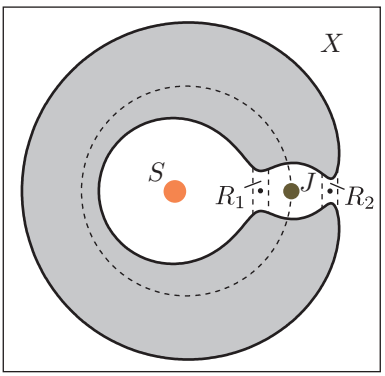


FIGURE 3.2.2. Division of Hill's region into five sets.

Let S and J be the large sets, i.e., realms, that contain the Sun and Jupiter, respectively; let region R_1 and region R_2 be those sets that contain

the two equilibrium points L_1 and L_2, respectively; and let X be the realm that lies exterior to the orbit of Jupiter. We also divide the energy surface $\mathcal{M}$ into sets projecting onto the sets shown in Figure 3.2.2. We will adopt the notation convention introduced in §2.4 where a set in the energy surface will be denoted by a script letter and its position space projection denoted by an italicized letter: e.g., region $\mathcal{R}_1$ for the set in the energy surface whose projection is the region R_1 in the position space. Theorem 3.2.1 leads to the assertion that *one can choose the division described above so that we simultaneously have sufficient control of the flow in both sets $\mathcal{S}$ and $\mathcal{R}_1$ to construct a homoclinic orbit. Theorem 3.2.2 makes the same assertion for sets $\mathcal{X}$ and $\mathcal{R}_2$.*

Let $\mathcal{R}$ denote either $\mathcal{R}_1$ or $\mathcal{R}_2$. As $\mathcal{R}$ is a function not only of b_1 and c_1, but also μ and e, we sometimes write it as $\mathcal{R}(\mu, e)$.

The analysis of $\mathcal{R}(\mu, e)$ is of a local nature. In fact, we limit ourselves to those energy values e for which the linearized equations about the equilibrium point give us the qualitative picture of the flow. The flow for the linearized equations was already analyzed in some detail in Chapter 2.

We know that for b_1 and c_1 close to 1, i.e., for the region $\mathcal{R}$ close to the periodic orbit (p.o.), the flow for $\mathcal{R}$ is that shown in Figure 2.9.1. But we also know that we cannot make c_1 arbitrarily large without disturbing this qualitative picture. On the other hand, we would like to make c_1 large enough to obtain accurate estimates on the behavior of the flow in $\mathcal{S}$ and $\mathcal{X}$. The following theorems show that there exists a c_1 which allows us to balance these two factors. In the theorem, $n_{i,j}$ is the n_j bounding sphere in region $\mathcal{R}_i$.

3.2.1 Theorem (McGehee)**.** *There exist constants b_1 and c_1 and an open set O_1 in the (μ, e)-plane* (see Figure 3.2.3(a)) *containing the graph of $e = E_1(\mu)$ for small $\mu > 0$ such that, for $(\mu, e) \in O_1$:*

1. *The energy surface $\mathcal{M}(\mu, e)$ contains an invariant torus $\mathcal{A}_1$ separating the Sun from Jupiter.*

2. *For $e > E_1(\mu)$, the flow in $\mathcal{R}_1(\mu, e)$ is qualitatively the same as the flow for the linearized equations.* (See Figure 2.9.1)

3. *If we let $\mathcal{T}_1$ be that submanifold of $\mathcal{M}$ co-bounded by the invariant torus $\mathcal{A}_1$ and the left bounding sphere $n_{1,1}$ of the equilibrium region $\mathcal{R}_1$* (see Figure 3.2.3(b))*, then there exists a function*

$$\theta : \mathcal{T}_1 \to \mathbb{R}$$

such that

   (a) *θ is a meridional angular coordinate for $\mathcal{T}_1$;*

   (b) *θ is strictly increasing along orbits.*

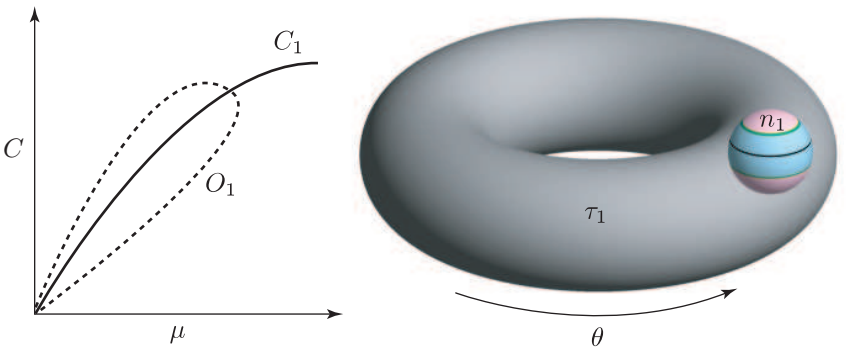


FIGURE 3.2.3. (a) Open set O_1 in (μ, e)-plane. (b) The region $\mathcal{T}_1$ with meridional angular coordinate θ.

3.2.2 Theorem (McGehee)**.** *There exist constants b_1 and c_1 and an open set O_2 in the (μ, e)-plane containing the graph of $e = E_2(\mu)$ for small $\mu > 0$ such that, for $(\mu, e) \in O_2$:*

1. *The energy surface $\mathcal{M}(\mu, e)$ contains an invariant torus $\mathcal{A}_2$ separating the Sun and Jupiter from infinity.*

2. *For $e > E_2(\mu)$, the flow in $\mathcal{R}_2(\mu, e)$ is qualitatively the same as the flow for the linearized equations. (See Figure 2.9.1)*

3. *If we let $\mathcal{T}_2$ be that submanifold of $\mathcal{M}$ co-bounded by the invariant torus $\mathcal{A}_2$ and the right bounding sphere $n_{2,2}$ of the equilibrium region $\mathcal{R}_2$, then there exists a function*

$$\theta : \mathcal{T}_2 \to \mathbb{R}$$

such that

   (a) *θ is a meridional angular coordinate for $\mathcal{T}_2$;*

   (b) *θ is strictly increasing along orbits.*

See McGehee [1969] for the proofs of these theorems.

Part 3 of the above theorems gives us the following properties for the flow in $\mathcal{T}_1$ and $\mathcal{T}_2$. The increase in θ along an orbit segment in $\mathcal{T}_1$ (or $\mathcal{T}_2$) with endpoints in the bounding sphere $n_{1,1}$ (or $n_{2,2}$, respectively) is close to a non-zero integer multiple of 2π. The increase in θ along any other orbit segment which can be deformed to the first, keeping both endpoints in the bounding sphere n, is close to the same integer multiple of 2π. Furthermore, the increase of θ along any orbit segment remaining for an arbitrarily long

time in $\mathcal{T}_1$ or $\mathcal{T}_2$ is arbitrary large. As will be shown, these are precisely the properties we need to carry out the proof of the existence of a homoclinic orbit.

The Existence of Orbits Homoclinic to a Lyapunov Periodic Orbit. Recall from Chapter 2 that for an energy $e > E_i$, there is a p.o. around L_i, $i = 1$ or 2, with two-dimensional invariant unstable, $W^u_{L_i,p.o.}$ and stable $W^s_{L_i,p.o.}$, manifolds, the asymptotic orbits which are locally diffeomorphic to cylinders. We recall that a homoclinic orbit related to a periodic orbit l is an orbit that tends to l as $t \to \pm\infty$. Therefore, it is on the stable and unstable invariant manifolds of l. A homoclinic orbit is called a ***transversal homoclinic orbit*** if at some point of the orbit the tangent spaces to the stable and unstable manifolds at that point span the full tangent space to $\mathcal{M}(\mu, e)$ at the same point.

We assert that in our problem *either a transversal homoclinic orbit exists, or "total degeneracy" occurs.* Total degeneracy is the case when every orbit asymptotic to the unstable periodic orbit at one end is also asymptotic to the same periodic orbit at the other end and hence is a homoclinic orbit. In other words, the total degeneracy situation occurs when the stable and unstable manifolds of the Lyapunov orbit coincide with each other. In either event we conclude the existence of a homoclinic orbit. We shall sketch the proof below for completeness. For more details, see Conley [1968] and McGehee [1969].

Assume that total degeneracy does not occur. The first step of the proof of the preceding assertion is to find an orbit segment in $\mathcal{T}_1$ connecting either d_1^- to a_1^+ or a_1^- to d_1^+ as follows. Consider Figure 3.2.4, where we show an example of the latter, and where n_1 denotes $n_{1,1}$. Since $\mathcal{T}_1$ is compact and our flow, which is Hamiltonian, preserves a nondegenerate area element, we can conclude that some orbit which leaves $\mathcal{R}_1$ (and crosses the bounding sphere n_1) and so enters $\mathcal{T}_1$ must also leave $\mathcal{T}_1$ and re-enter $\mathcal{R}_1$ (and recross n_1). Therefore, for some point $p \in d_1^-$ of n_1, there is an orbit segment connecting p to a point $q \in d_1^+$ of n_1. Recall from §2.7 and Figure 2.7.2 that in $\mathcal{R}_1$, the spherical caps d_1^- and d_1^+ are where the flow crosses n_1.

Starting with this orbit segment connecting p to q, we can find an orbit segment connecting either d_1^- to a_1^+ or a_1^- to d_1^+ as follows. Let γ be an arc in d_1^- linking p to a_1^- (where $\gamma \cap a_1^-$ is not on a homoclinic orbit). If all of γ is carried by the flow to the spherical cap d_1^+, then we shall have an orbit segment with one endpoint in a_1^- and the other in d_1^+. Otherwise, starting from p, there is some maximal initial half-open subarc γ' of γ which is carried by the flow to d_1^+. Let r be the first point of γ not in γ', then the orbit segment with one endpoint at r must become arbitrarily long. But the only way this orbit segment can become arbitrarily long is to approach the asymptotic set, since the number of times it can wind around $\mathcal{T}_1$ is finite and therefore must contain an arbitrarily long subsegment in $\mathcal{R}_1$. Because of our knowledge of the flow in $\mathcal{R}_1$, we know that long orbit segments in

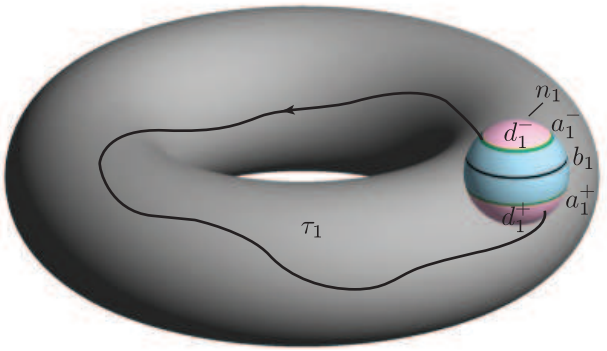


FIGURE 3.2.4. The existence of orbits homoclinic to the Lyapunov orbit. The trajectory shown connects a_1^- to d_1^+.

$\mathcal{R}_1$ must lie close to the cylinders of asymptotic orbits and therefore r must be carried to a_1^+. Hence, in either case we conclude that there is an orbit segment connecting the set $d_1^\pm$ in one hemisphere to the set of asymptotic orbits in the other.

Without loss of generality, we can suppose that we have found an orbit segment with one endpoint, called α, in a_1^- and the other in d_1^+. We now choose for γ the whole set a_1^-. Using arguments similar to the above, we can conclude that either all of a_1^- is carried by the flow inside d_1^+, or there exists a point $\beta \in a_1^-$ such that the orbit segment with β as an endpoint becomes asymptotic at the other end. If the first possibility holds, we would have a map of d_1^- to the interior of d_1^+, contradicting area preservation of Hamiltonian flow. Thus we have proven that either transversal homoclinic orbits exist or total degeneracy occurs for the interior realm. The same proof also works for the exterior realm.

3.3 Existence of Transversal Homoclinic Orbits in the Interior Realm

Conley [1968] and McGehee [1969] did not settle the issue of when one has transversality of the homoclinic orbit families for the PCR3BP, since total degeneracy was a possibility they could not rule out analytically. Subsequently, Llibre, Martinez, and Simó [1985] devoted their major effort to show that under appropriate conditions, the invariant manifolds of the L_1 Lyapunov orbits do meet transversely. In this section, we summarize their

analytical results. Moreover, in §3.5 we explore numerically the existence of transversal homoclinic orbits in both the interior and exterior realms.

To state the major analytical results of Llibre, Martinez, and Simó [1985], we first need to set up some notation. As mentioned earlier, near L_1 and for energy values e where $E_1 < e < E_2$ (case 2) there is a family of unstable Lyapunov orbits. When e approaches E_1 from above, the periodic orbit tends to L_1. There are one-dimensional invariant stable, $W^s_{L_1}$, and unstable, $W^u_{L_1}$, manifolds associated to L_1.

Notice that equations (2.5.1) have the following symmetry

$$s : (x, y, v_x, v_y, t) \to (x, -y, -v_x, v_y, -t). \tag{3.3.1}$$

Therefore, if we know the unstable manifold of L_1 of the Lyapunov orbit (which is a symmetrical periodic orbit) the corresponding stable manifold is obtained through the use of the stated symmetry. This observation will be used in later sections to find the transversal homoclinic orbits.

Analytical Results for L_1 Lyapunov Orbit in Interior Realm. Using the basic framework developed in McGehee [1969], Llibre, Martinez, and Simó [1985] were able to prove the following two theorems. Together these two theorems imply that for sufficiently small μ and for an appropriate range of $\Delta E = e - E_1$, the invariant manifolds $W^{s,\mathcal{S}}_{L_1,\text{p.o.}}$ and $W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$ in the interior realm $\mathcal{S}$ intersect transversely.

3.3.1 Theorem (Llibre-Martinez-Simó)**.** *For μ sufficiently small, the branch $W^{u,\mathcal{S}}_{L_1}$ of $W^u_{L_1}$ in the interior realm $\mathcal{S}$ has a projection on position space (see Figure 3.3.1(a)) given by*

$$\begin{aligned} d &= \mu^{1/3}\left(\frac{2}{3}N - 3^{1/6} + M\cos t + o(1)\right), \\ \alpha &= -\pi + \mu^{1/3}(Nt + 2M\sin t + o(1)), \end{aligned}$$

where d is the distance to the zero velocity curve, α is the angular coordinate and N and M are constants.

In particular, for a sequence of values of μ which have the following asymptotic expression:

$$\mu_k = \frac{1}{N^3 k^3}(1 + o(1)), \tag{3.3.2}$$

the first intersection of this projection with the x-axis is orthogonal to that axis, giving a symmetric (1,1)-homoclinic orbit for L_1. The prefix (1,1) refers to the first intersection (with the Poincaré section defined by the plane $y = 0, x < 0$) of both the stable and unstable manifolds of L_1.

3.3.2 Theorem (Llibre-Martinez-Simó)**.** *For μ and $\Delta E = e - E_1$ sufficiently small, the branch $W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$ of $W^u_{L_1,\text{p.o.}}$ contained initially in the*

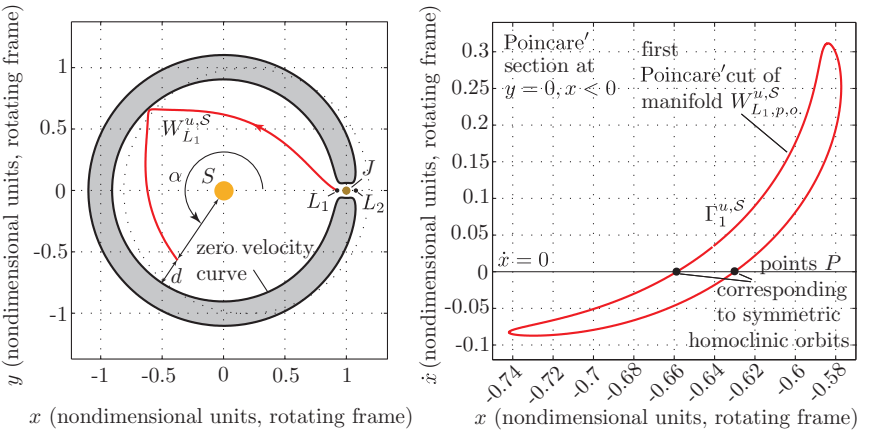


FIGURE 3.3.1. (a) Projection of the interior branch of the manifold $W^u_{L_1}$ on the position space. (b) First intersection (Poincaré "cut") $\Gamma^{u,\mathcal{S}}_1$ of the interior branch of $W^u_{L_1,\text{p.o.}}$ with the plane $y = 0, x < 0$.

interior realm $\mathcal{S}$ of the energy surface intersects the plane $y = 0$ for $x < 0$ in a curve diffeomorphic to a circle (see Figure 3.3.1(b)).

In particular, for points in the (μ, e) plane such that there is a μ_k of Theorem 3.3.1 for which

$$\Delta E > L\mu_k^{4/3}(\mu - \mu_k)^2 \tag{3.3.3}$$

holds (where L is a constant), there exist symmetric transversal (1,1)-homoclinic orbits.

For details of the proofs, see Llibre, Martinez, and Simó [1985]. We would like to make a few comments about these results which are pertinent to the main thrust of this chapter.

1. The main objective of both theorems (3.3.1 and 3.3.2) is to study the transversality of the invariant manifolds for the L_1 Lyapunov orbit on the energy surface whose energy e is slightly greater than $E_1(\mu)$ as one varies μ and e. The main step is to obtain an expression for the first intersection of the unstable manifold $W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$ with the plane $y = 0$ in the region $x < 0$, which we label $\Gamma^{u,\mathcal{S}}_1$. While formulas were provided in Llibre, Martinez, and Simó [1985] for this closed curve as a function of μ and ΔE in the variables

$x, \dot{x}$, they are quite complicated and difficult to interpret and hence are not included here. But the key point is the following. According to Theorem 3.3.1, the set of values of μ for which we have a symmetric (1,1)-homoclinic orbit associated to L_1 is discrete and is given by equation (3.3.2). Then for any other value of μ the unstable manifold $W^{u,\mathcal{S}}_{L_1}$ of L_1 reaches the $(x, \dot{x})$-plane in a point $(x_1, \dot{x}_1)$ outside $\dot{x} = 0$. Therefore, if ΔE is too small, $\Gamma^{u,\mathcal{S}}_1$ does not cut the x-axis and hence (by symmetry), the intersection $\Gamma^{s,\mathcal{S}}_1$ of the stable manifold $W^{s,\mathcal{S}}_{L_1,\text{p.o.}}$ with the plane $y = 0, x < 0$ does not cut the x-axis either. Therefore the first intersections of the invariant manifolds do not meet and there is no symmetric (1,1)-homoclinic orbit.

However, for a fixed value of μ, if we increase ΔE, we hope that $\Gamma^{u,\mathcal{S}}_1$ of the unstable manifold will become large. Therefore we can look for some value of ΔE such that $\Gamma^{u,\mathcal{S}}_1$ becomes tangent to the x-axis or even intersects it at more than one point. Then, due to the symmetry of the PCR3BP (3.3.1), $\Gamma^{s,\mathcal{S}}_1$ of the stable manifold also intersects the x-axis at the same points. Points P on the x-axis where $\Gamma^{u,\mathcal{S}}_1$ and $\Gamma^{s,\mathcal{S}}_1$ intersect correspond to (symmetric) orbits homoclinic to the Lyapunov orbit (see Figure 3.3.1(b)). If $\Gamma^{u,\mathcal{S}}_1$ is transversal to $\Gamma^{s,\mathcal{S}}_1$ at P then the homoclinic orbit is transversal. The results of Theorem 3.3.2 say that the above phenomenon occurs if $\Delta E > L\mu_k^{4/3}(\mu - \mu_k)^2$ holds.

2. Using the results of Theorem 3.3.2, Llibre, Martinez, and Simó [1985] were able to draw the mesh of homoclinic tangencies for the $(\mu, \Delta E)$-plane. The numbers in Figure 3.3.2 show the number of symmetric (1,1)-homoclinic points found in the first intersection of $W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$ with the plane

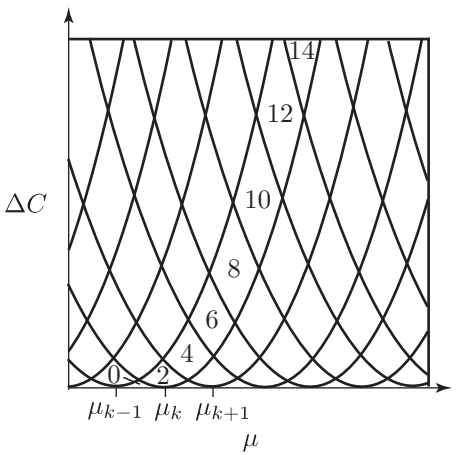


FIGURE 3.3.2. Partition of the $(\mu, \Delta E)$-plane according to the number of symmetric (1,1)-homoclinic points found in the first intersection of $W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$ with the plane $y = 0, x < 0$.

$y = 0, x < 0$ when one varies μ and ΔE. For us, the key point of the theorems is that for the wide range of μ which exist in the solar system, the invariant manifolds of the L_1 Lyapunov orbit intersect transversely for sufficiently large ΔE.

3. The heart of the proofs of these two theorems is to obtain expressions for $W^{u,\mathcal{S}}_{L_1}$ as a function of μ and for $W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$ as a function of μ and ΔE. By using the basic framework of McGehee [1969], Llibre, Martinez, and Simó [1985] divided the annulus T_1 in the interior realm S into two sets: a small neighborhood H near R_1 and the rest of the realm outside this small neighborhood. In the neighborhood H, the PCR3BP can be considered as a perturbation of the Hill's problem. In celestial mechanics, it is well known that Hill's problem studies the behavior near the small mass of the PCR3BP in the limit when μ approaches zero. In the rest of the realm away from the small mass, the PCR3BP can be approximated by the two-body problem in a rotating frame. Through a number of careful estimations, Llibre, Martinez, and Simó [1985] were able to obtain these analytical results.

Summary. Conley [1968] and McGehee [1969] proved the existence of homoclinic orbits for both the interior and exterior realm, and Llibre, Martinez, and Simó [1985] showed analytically the existence of transversal symmetric (1,1)-homoclinic orbits in the interior realm under appropriate conditions. For our problem, we need to find transversal homoclinic orbits in both interior and exterior realms as well as transversal heteroclinic cycles for the L_1 and L_2 Lyapunov orbits. In the following sections, we perform some numerical explorations using the methods described in Chapter 4. For more details on finding invariant manifolds numerically, see Gómez, Jorba, Masdemont, and Simó [1991a] and references therein.

3.4 Existence of Transversal Homoclinic Orbits in the Exterior Realm

We turn our attention now to numerical explorations of the problem, and in particular, to the existence of transversal homoclinic orbits for the L_2 Lyapunov orbit in the exterior realm. Though there are no analytical results proving the existence of transversal homoclinic orbits in the $\mathcal{X}$ realm, we can construct them numerically by finding an intersection of the manifolds $W^s_{L_2,\text{p.o.}}$ and $W^u_{L_2,\text{p.o.}}$ on an appropriately chosen Poincaré section.

Numerical experiments guided by geometrical insight suggest that we cut the flow by the plane $y = 0$, the line passing through the two masses in the rotating frame. The branch of the manifold $W^u_{L_2,\text{p.o.}}$ which enters the $\mathcal{X}$ realm flows clockwise in the position space, as shown in See Figure 3.4.1(a). We refer to this exterior branch of the manifold as $W^{u,\mathcal{X}}_{L_2,\text{p.o.}}$. Outside of a neighborhood of n_2 in the $\mathcal{X}$ realm, this two-dimensional manifold

tube $W^{u,\mathcal{X}}_{L_2,\text{p.o.}}$ first intersects the plane $y = 0$ on the part of $\mathcal{T}_2$ which is opposite to L_2 with respect to the Sun (i.e., $x < 0$). The intersection shown in Figure 3.4.1(b) is a curve diffeomorphic to a circle, as one would expect geometrically. We call this intersection the first *cut* of the tube $W^{u,\mathcal{X}}_{L_2,\text{p.o.}}$ with $y = 0$. Some arcs of this curve produce successive intersections with-

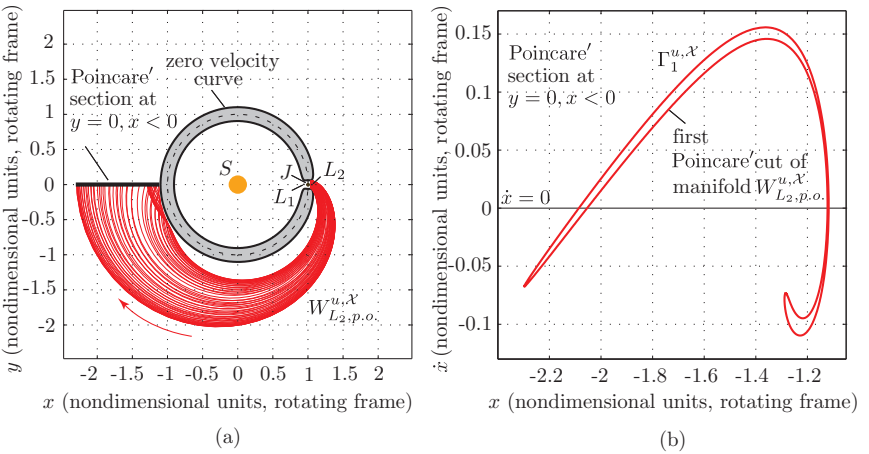


FIGURE 3.4.1. (a) The position space projection of the unstable manifold "tube" $W^{u,\mathcal{X}}_{L_2,\text{p.o.}}$ until the first intersection with the Poincaré section at $y = 0, x < 0$. (b) The first Poincaré cut $\Gamma^{u,\mathcal{X}}_1$ of the manifold $W^{u,\mathcal{X}}_{L_2,\text{p.o.}}$ on the plane $y = 0, x < 0$.

out leaving the $\mathcal{X}$ realm. The q-th of these intersections of $W^{u,\mathcal{X}}_{L_2,\text{p.o.}}$ with $y = 0$ will be referred to as $\Gamma^{u,\mathcal{X}}_q$. In a similar manner we call $\Gamma^{s,\mathcal{X}}_p$ the corresponding p-th intersection with $y = 0$ of $W^{s,\mathcal{X}}_{L_2,\text{p.o.}}$.

A point in $y = 0$ belonging to $\Gamma^{u,\mathcal{X}}_q \cap \Gamma^{s,\mathcal{X}}_p$ (if not empty) will be called a (q,p)-***homoclinic point***. The existence of (q,p)-homoclinic points for certain q and p is shown in McGehee [1969].

Our goal is to obtain the first such transversal intersection of $\Gamma^{u,\mathcal{X}}_q$ with $\Gamma^{s,\mathcal{X}}_p$ and so obtain a *transversal* (q,p)*-homoclinic point*. The (q,p)-homoclinic point P is transversal if $\Gamma^{u,\mathcal{X}}_q$ and $\Gamma^{s,\mathcal{X}}_p$, which necessarily intersect at P, do so transversely: that is, their tangent spaces span the $(x,\dot{x})$-plane at P. Other intersections (for larger q and p) may exist, but we will restrict ourselves for now to the first. Suppose that the unstable manifold intersection $\Gamma^{u,\mathcal{X}}_q$ is a closed curve γ in the variables $x, \dot{x}$. Let s_x be the symmetry with respect to the x-axis on this plane. Then due to the symmetry of the PCR3BP (3.3.1), the q-th intersection $\Gamma^{s,\mathcal{X}}_q$ of the stable manifold $W^{s,\mathcal{X}}_{L_2,\text{p.o.}}$ with $y = 0$ is $s_x\gamma$. For some minimum q, the closed curve γ intersects the $\dot{x} = 0$ line of the $(x,\dot{x})$-plane. Points P along the curve γ which intersect the $\dot{x} = 0$ line are (q,q)-homoclinic points, corresponding

to (symmetric) orbits homoclinic to the Lyapunov orbit. If the curve γ is transversal to the curve $s_x\gamma$ at the point P then the homoclinic orbit corresponding to P is transversal. If intersections between the curves γ and $s_x\gamma$ exist off the line $\dot{x} = 0$ (i.e., if the set $(\gamma \cap s_x\gamma)\backslash\{\dot{x} = 0\}$ is nonempty), then nonsymmetric homoclinic orbits exist.

Consider Figure 3.4.1(b), where we use the values $\mu = 9.537 \times 10^{-4}$ and $\Delta E = e - E_2 = 0.005$ to compute the unstable Poincaré cut $\Gamma_1^{u,\mathcal{X}}$. If we also plot the stable cut $\Gamma_1^{s,\mathcal{X}}$, which is the mirror image of unstable cut $\Gamma_1^{u,\mathcal{X}}$ (i.e., $s_x\Gamma_1^{s,\mathcal{X}}$), we find several points of intersection. In Figure 3.4.2(a),

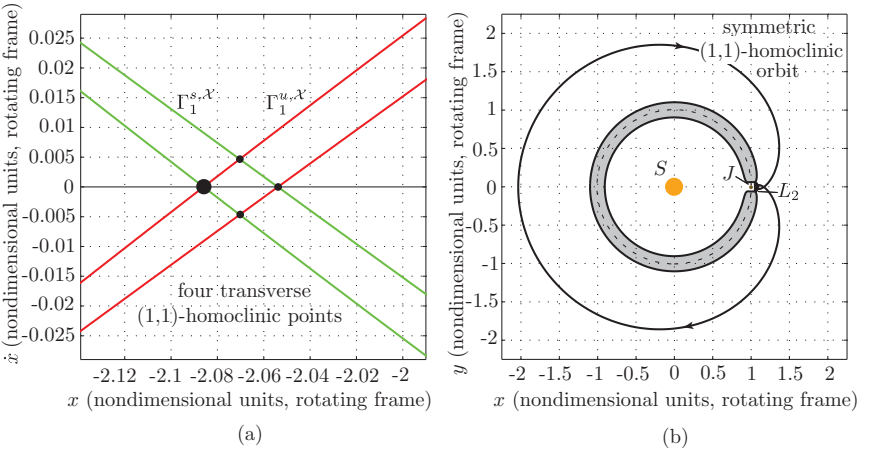


FIGURE 3.4.2. (a) A group of four transverse $(1, 1)$-homoclinic points. (b) The symmetric $(1, 1)$-homoclinic orbit corresponding to the left $\dot{x} = 0$ $(1, 1)$-homoclinic point (the large black dot in (a)).

we focus on the left-most group of points, centered at about $x = -2.07$. We find two $\dot{x} = 0$ intersections which are transversal homoclinic points in the $\mathcal{X}$ realm. The transversal symmetric $(1, 1)$-homoclinic orbit corresponding to the left $\dot{x} = 0$ intersection is shown in Figure 3.4.2(b).

We also notice two off-axis intersections in Figure 3.4.2(a), completing the local transversal intersection of two closed loops in the $(x, \dot{x})$-plane. As these two intersections occur near the line $\dot{x} = 0$, the appearance of the corresponding homoclinic orbits in position space will be nearly symmetric. A more pronounced case of nonsymmetry occurs for the other group of intersection points centered near $x = -1.15$ on the right side of Figure 3.4.1(b), for which we have the nonsymmetric $(1, 1)$-homoclinic orbit given in Figure 3.4.3.

Homoclinic Orbits in the Exterior and Jupiter Realms. A similar procedure can numerically produce homoclinic orbits in the interior realm as well as in the Jupiter realm. We can even look at cuts beyond the first and

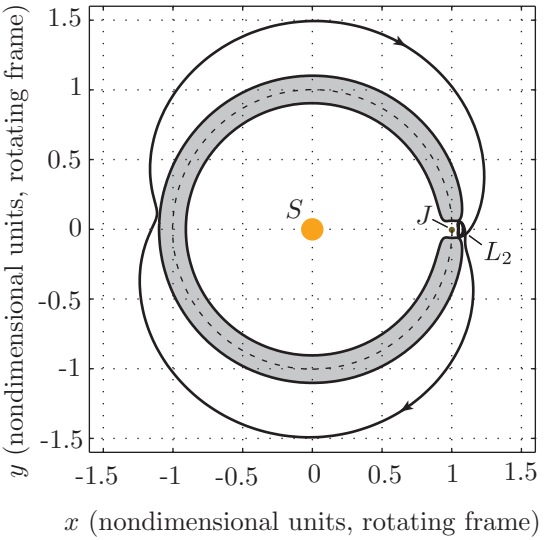


FIGURE 3.4.3. A nonsymmetric $(1,1)$-homoclinic orbit.

large values of μ and ΔE, such as shown in Figure 3.4.4(a). For example,

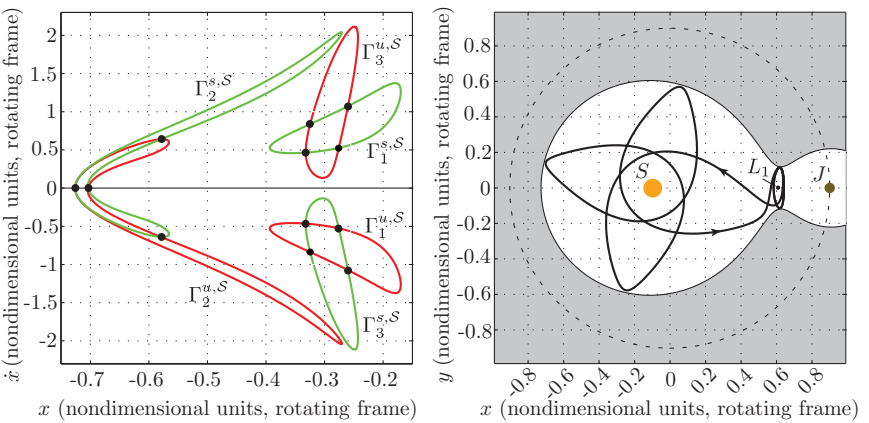


FIGURE 3.4.4. (a) The first three Poincaré cuts of the unstable ($W^{u,\mathcal{S}}_{L_1,\text{p.o.}}$) and stable ($W^{s,\mathcal{S}}_{L_1,\text{p.o.}}$) manifolds with the plane $y=0$. (b) A nonsymmetric $(1,3)$-homoclinic orbit in the interior realm (corresponding to the three large dots in (a)).

in Figure 3.4.4(b) we show an interior realm $(1,3)$-homoclinic orbit (note, also $(2,2)$ and $(3,1)$, using $\bar{q}+\bar{p}=q+p$) associated to an L_1 Lyapunov orbit for $\mu=0.1, \Delta E = e - E_1 = 0.03715$.

3.5 Existence of Heteroclinic Connections between Lyapunov Orbits

We construct a heteroclinic connection between Lyapunov orbits of L_1 and L_2 by finding an intersection of their respective invariant manifolds in the $\mathcal{J}$ realm. To do so, we seek points of intersection on a suitably chosen Poincaré section. For instance, to generate a heteroclinic orbit which goes from an L_1 Lyapunov orbit (as $t \to -\infty$) to an L_2 Lyapunov orbit (as $t \to +\infty$), we proceed as follows.

We restrict ourselves for now to case 3 ($e \in (E_2, E_3)$; see Figure 2.4.2), for which the Hill's region opens enough to permit Lyapunov orbits about both L_1 and L_2 to exist. Let the branch of the unstable manifold of the L_1 Lyapunov orbit which enters the $\mathcal{J}$ realm be denoted $W^{u,\mathcal{J}}_{L_1,\text{p.o.}}$. On the same energy surface there is an L_2 Lyapunov orbit, whose stable manifold in the $\mathcal{J}$ realm we similarly denote $W^{s,\mathcal{J}}_{L_2,\text{p.o.}}$. The projection of the two-dimensional manifold tubes onto the position space is shown in Figure 3.5.1(a).

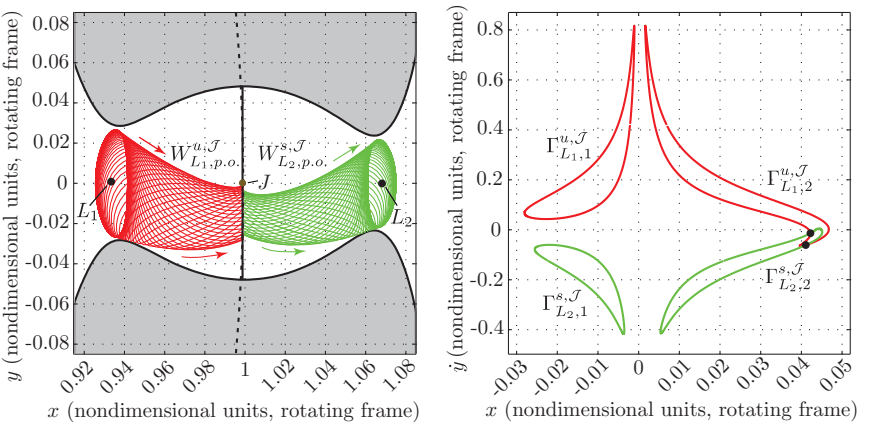


FIGURE 3.5.1. (a) The projection of invariant manifolds $W^{u,\mathcal{J}}_{L_1,\text{p.o.}}$ and $W^{s,\mathcal{J}}_{L_2,\text{p.o.}}$ in the realm J of the position space. (b) The first two Poincaré cuts of the invariant manifolds with the plane $x = 1 - \mu$.

To find intersections between these two tubes, we cut the flow by the plane $x = 1 - \mu$, denoted by the thick black line in Figure 3.5.1(a). The cuts on this plane are shown in Figure 3.5.1(b).

This convenient plane maximizes the number of intersections for values of μ, e which produce manifolds making a limited number of revolutions around Jupiter before escaping from the $\mathcal{J}$ realm. The q-th intersection of

$W^{u,\mathcal{J}}_{L_1,\text{p.o.}}$ with the plane $x = 1-\mu$ will be labeled $\Gamma^{u,\mathcal{J}}_{L_1,q}$. Similarly, we will call $\Gamma^{s,\mathcal{J}}_{L_2,p}$ the p-th intersection of $W^{s,\mathcal{J}}_{L_2,\text{p.o.}}$ with $x = 1-\mu$.

Numerical experiments show that the L_1 Lyapunov orbit unstable manifold $W^{u,\mathcal{J}}_{L_1,\text{p.o.}}$ does not coincide with the L_2 Lyapunov orbit stable manifold $W^{s,\mathcal{J}}_{L_2,\text{p.o.}}$. Moreover, for a wide range of μ and e values (where $e \in (E_2, E_3)$), numerical explorations demonstrate that they do intersect transversely. While we recognize that for certain values of μ and e, there are tangencies between the stable and unstable manifold, we will not deal with this interesting case in this book. Hence, from now on, we will concentrate our numerical explorations only on the cases where the stable and unstable manifold intersect transversely.

Now, suppose that $\Gamma^{u,\mathcal{J}}_{L_1,q}$ and $\Gamma^{s,\mathcal{J}}_{L_2,p}$ are each closed curves in the variables $y, \dot{y}$. A point in the plane $x = 1-\mu$ belonging to the intersection of the two closed curves (i.e., $\Gamma^{u,\mathcal{J}}_{L_1,q} \cap \Gamma^{s,\mathcal{J}}_{L_2,p}$) will be called a (q,p)-***heteroclinic point*** because such a point corresponds to a heteroclinic orbit going from the L_1 Lyapunov orbit to the L_2 Lyapunov orbit. Moreover, since we restrict ourselves to the case where $W^{u,\mathcal{J}}_{L_1,\text{p.o.}}$ and $W^{s,\mathcal{J}}_{L_2,\text{p.o.}}$ intersect transversely, the (q,p)-heteroclinic point will be a *transversal heteroclinic point*. Our objective is to obtain the first intersection point (or group of points) of the curve $\Gamma^{u,\mathcal{J}}_{L_1,q}$ with the curve $\Gamma^{s,\mathcal{J}}_{L_2,p}$ and so obtain the minimum values of q and p such that we have a transversal (q,p)-heteroclinic point. Other intersections may exist, but we will restrict ourselves for now to the first. For some minimum q and p, we have an intersection of the curves, and some number of (q,p)-heteroclinic points, depending on the geometry of the intersection. Note that the sum $q+p$ must be an even positive integer.

As we are interested in heteroclinic points for the Sun-Jupiter system ($\mu = 9.537 \times 10^{-4}$), we take $e = -1.5185$ and numerically obtain the intersections of the invariant manifolds $W^{u,\mathcal{J}}_{L_1,\text{p.o.}}$ and $W^{s,\mathcal{J}}_{L_2,\text{p.o.}}$ with the plane $x = 1-\mu$. In Figure 3.5.1(b) we show the curves $\Gamma^{u,\mathcal{J}}_{L_1,q}$ for $q = 1, 2$ and $\Gamma^{s,\mathcal{J}}_{L_2,p}$ for $p = 1, 2$. Notice that $\Gamma^{u,\mathcal{J}}_{L_1,2}$ and $\Gamma^{s,\mathcal{J}}_{L_2,2}$ intersect in two points (the black dots in Figure 3.5.1(b) near $y = 0.042$). Thus, the minimum q and p for a heteroclinic point to appear for these particular values of μ, e are $q = 2$ and $p = 2$. The $(2,2)$-heteroclinic points can each be forward and backward integrated to produce heteroclinic orbits going from the L_1 Lyapunov orbit to the L_2 Lyapunov orbit, otherwise known as a *heteroclinic connection*. We show one of the heteroclinic orbits in Figure 3.5.2. Notice that the number of revolutions around Jupiter is given by $(q+p-1)/2$. The reverse orbit, going from the L_2 Lyapunov orbit to the L_1 Lyapunov orbit, is easily given by the symmetry s (3.3.1). It is the mirror image (about the x-axis) of the trajectory in Figure 3.5.2, with the direction arrows reversed. These two heteroclinic connections together form a symmetric ***heteroclinic cycle***.

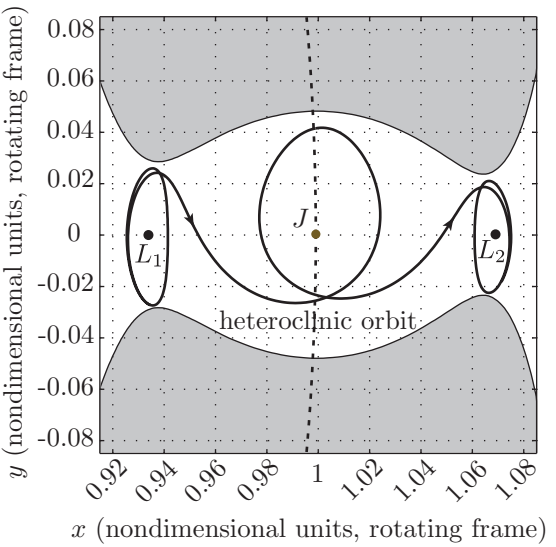


FIGURE 3.5.2. The existence of a transversal $(2, 2)$-heteroclinic orbit in the J realm.

3.6 Existence of Chains of Homoclinic Orbits and Heteroclinic Cycles

We have numerically demonstrated the existence of homoclinic and heteroclinic orbits associated to the L_1 and L_2 Lyapunov orbits for case 3. We now take the final step, combining homoclinic and heteroclinic orbits of the same energy value to generate what is called a ***homoclinic-heteroclinic chain*** of orbits, which connect asymptotically the L_1 and L_2 Lyapunov orbits to each other. As will be seen, these chains imply a complicated dynamics connecting the interior, exterior, and Jupiter realms.

As an example, we again choose the Sun-Jupiter system ($\mu = 9.537 \times 10^{-4}$), but now an energy value similar to that of comet Oterma during its Jupiter encounters ($e = -1.515$). Using the described methodologies, we obtain an interior realm orbit homoclinic to the L_1 Lyapunov orbit, an exterior realm orbit homoclinic to the L_2 Lyapunov orbit, and a heteroclinic cycle connecting the L_1 and L_2 Lyapunov orbits in the Jupiter realm. The union of these orbits is the homoclinic-heteroclinic chain shown in Figure 3.1.1. The existence of homoclinic-heteroclinic chains has important consequences, which will be expanded upon further in the following sections.

3.7 Construction of a Suitable Poincaré Map

The idea of reducing the study of the global orbit structure of a system of differential equations to the study of an associated discrete map is due to Poincaré [1890], who first utilized the method in his studies of the restricted three-body problem. In this section we use the chain of two homoclinic orbits and one symmetric heteroclinic cycle (such as the one shown in Figure 3.1.1) to construct a suitable Poincaré map. Our choice of Poincaré map will allow us to study the complex global orbit structure near the chain. We find an invariant set for this map near some transversal homoclinic and heteroclinic points along the chain where "Smale horseshoe"-like dynamics exist. We then use symbolic dynamics to characterize the chaotic motion of a comet in a neighborhood of the chain as it transitions intermittently through the interior, Jupiter and exterior realms. Not only do we prove the existence of the invariant set, but we also numerically approximate it in Chapter 4, gaining further insight into the complex global dynamics associated with the chains.

Here is additional detail about how we proceed: In this section, we construct a Poincaré map P transversal to the flow around a chain whose domain U consists of four different squares $U_i, i = 1, 2, 3, 4$, located in different parts of phase space in the neighborhood of the chain, as shown schematically in Figure 3.7.1.

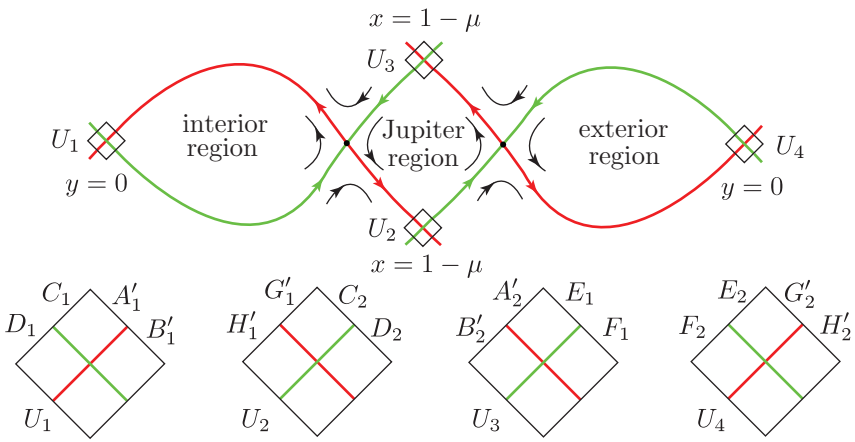


FIGURE 3.7.1. The construction of a suitable Poincaré map. The labeling D_1, etc., is described in the text.

Squares U_1 and U_4 are contained in the surface $y = 0$ and each centers around a transversal homoclinic point in the interior and the exterior realm,

respectively. Squares U_2 and U_3 are contained in the surface $x = 1 - \mu$ ($y < 0$ and $y > 0$, respectively) and center around transversal heteroclinic points in the Jupiter realm which are symmetric with respect to each other. Clearly, for any orbit which passes through a point q in one of the squares and whose images and pre-images $(P^n(q), n = 0, \pm 1, \pm 2, \ldots)$ all remain in the domain U, the whereabouts of $P^n(q)$ (as n increases or decreases) can provide some of the essential information about the history of the particular orbit. We record this history with a bi-infinite sequence. This well-known technique of studying only the set of points that forever remain in the domain U (the ***invariant set***) provides us with all the periodic solutions as well as the recurrent solutions in the neighborhood of the chain.

The technique of characterizing the orbit structure of a dynamical system via a set of bi-infinite sequences of "symbols" is known as ***symbolic dynamics***.

In §3.8 and §3.9, we extend the symbolic dynamics results of Llibre, Martinez, and Simó [1985] to our situation and construct a set of bi-infinite sequences with two families of symbols. The first family is a *subshift of finite type* with four symbols $\{u_1, u_2, u_3, u_4\}$. It is used to keep track of the whereabouts of an orbit with respect to the four squares U_1, U_2, U_3, U_4. The symbol u_i is recorded every time the U_i square is pierced by the orbit. Subshift here means that among the set of all bi-infinite sequences of four symbols, (i.e., $(\ldots, u_{i_{-1}}; u_{i_0}, u_{i_1}, u_{i_2}, \ldots)$ where i_j ranges from 1 to 4), certain sequences where the adjacent entries in the sequence violate certain relations are not allowed. For example, from U_1, the (forward) flow cannot get to U_4 without passing through other squares. Hence, in the bi-infinite sequence, the symbol u_1 cannot be followed by u_4. The relations can be defined by a matrix A called the *transition matrix*. In our case,

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

It is constructed by the following rule: $(A)_{kl} = 1$ if the ordered pair of symbols u_k, u_l may appear as adjacent entries in the symbolic sequence, and $(A)_{kl} = 0$ if the ordered pair of symbols u_k, u_l may not appear as adjacent entries. For example, since u_1 cannot be followed by u_4, we have $(A)_{14} = 0$.

The second family is a *full shift of infinite type* with symbols of positive integers greater than a fixed integer m. This set of bi-infinite sequences of positive integers is used to keep track of the number of integer revolutions that the projection of an orbit winds around either L_1 or L_2 when the orbit enters the equilibrium regions $\mathcal{R}_1$ or $\mathcal{R}_2$, respectively.

In §3.10, we state the main theorem of this chaoter and discuss its implications. The theorem gives the global orbit structure of the PCR3BP in a neighborhood of a chain of homoclinic orbits and a symmetric heteroclinic

cycle. It says essentially that given any bi-infinite sequence

$$\alpha = (u, r) = (\ldots, (u_{i_{-1}}, r_{-1}); (u_{i_0}, r_0), (u_{i_1}, r_1), (u_{i_2}, r_2) \ldots),$$

there exist initial conditions near the transversal homoclinic and heteroclinic points (the intersection of the chain with U) such that an orbit corresponding to such initial conditions starts at U_{i_0} and goes to U_{i_1} (provided $(A)_{i_0 i_1} = 1$). This orbit passes through either the equilibrium region $\mathcal{R}_1$ or $\mathcal{R}_2$ depending on whether the initial index (i_0 in the current case) is $1, 3$ or $2, 4$ (see Figure 3.7.1 for reference). For example if $i_0 = 1$, then the projection of the orbit winds around L_1 for r_0 revolutions inside the region $\mathcal{R}_1$ before leaving for U_{i_1}. After that, the same process begins with (u_{i_1}, r_1) in place of (u_{i_0}, r_0) and (u_{i_2}, r_2) in place of (u_{i_1}, r_1), etc. For negative time, a similar behavior is described for $(u_{i_{-1}}, r_{-1}), (u_{i_0}, r_0)$, etc. While the formalism involved in the proof is fairly standard, there are a few new features which may be worth pointing out. While most of these comments will be made earlier, we provide a sketch of the proof in §3.10 both for completeness and for the convenience of the reader. For more details, one can consult Koon, Lo, Marsden, and Ross [2000], Moser [1973], Llibre, Martinez, and Simó [1985], Wiggins [2003], and Wiggins [1993].

In Chapter 4 we numerically construct sets of orbits with prescribed itineraries. By successive application of the Poincaré map P to a transversal plane in a (rather large) neighborhood of a chain, we can theoretically generate regions of orbits with itineraries of any size.

Construction of a Suitable Poincaré Map. In §3.6, we showed that with an appropriate energy value, there exists a chain of two homoclinic orbits and one symmetric heteroclinic cycle. For simplicity of exposition, let us suppose that the chain $\mathcal{C}$ consists of $(1, 1)$-transversal homoclinic orbits in the interior and exterior realms and a symmetric $(1, 1)$-transversal heteroclinic cycle in the Jupiter realm. A similar study can be done for other cases.

Now we are ready to construct a Poincaré map. The first step is to construct the transversal maps on the bounding spheres of the equilibrium regions $\mathcal{R}_1$ and $\mathcal{R}_2$. Let ϵ_1 and ϵ_2 be small positive quantities. For the bounding spheres $n_{1,1}$ and $n_{1,2}$ of the equilibrium region $\mathcal{R}_1$, we define $\{A_1, B_1, C_1, D_1\}$, $\{E_1, F_1, G_1, H_1\}$ as the set of points of $\{d^-_{1,1}, r^-_{1,1}, r^+_{1,1}, d^+_{1,1}\}$, $\{d^+_{1,2}, r^+_{1,2}, r^-_{1,2}, d^-_{1,2}\}$, respectively, such that $||\zeta|^2 - \rho^*| < \epsilon_1$. These sets correspond to thin strips on the bounding sphere centered on the asymptotic sets $\{a^-_{1,1}, a^+_{1,1}\}$, $\{a^+_{1,2}, a^-_{1,2}\}$, respectively, as shown in Figure 3.7.2. Similarly, given ϵ_2 small, we can define corresponding strips for the bounding spheres $n_{2,1}$ and $n_{2,2}$ of the equilibrium region $\mathcal{R}_2$.

If ϵ_1 and ϵ_2 are small enough, the flow is transversal to the surfaces just defined. Recall from §2.8 that orbits entering $\mathcal{R}_1$ through C_1, D_1, E_1, F_1 leave it through B_1, H_1, A_1, G_1, respectively, because $|\zeta|^2$ is a first integral in $\mathcal{R}_1$. Therefore the diffeomorphisms $\psi_{1,i}$ send D_1, E_1, C_1, F_1 into

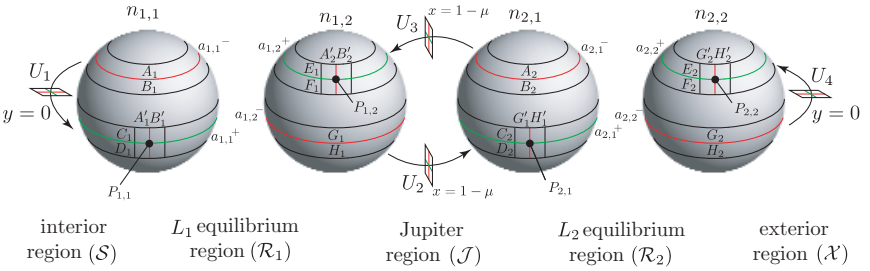


FIGURE 3.7.2. The strips near the asymptotic sets on the spheres $n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}$.

H_1, A_1, B_1, G_1 respectively, for $i = 1, 2, 3, 4$, where the $\psi_{1,i}$ are the mappings given in (2.8.1) and (2.8.2). Similar results hold for orbits entering $\mathcal{R}_2$ and the corresponding diffeomorphisms $\psi_{2,i}$ send D_2, E_2, C_2, F_2 into H_2, A_2, B_2, G_2 respectively, for $i = 1, 2, 3, 4$.

The second step is to construct transversal maps outside of the equilibrium regions. Let $p_{1,1} \in a_{1,1}^+$ (resp. $p_{2,2} \in a_{2,2}^+$) be a point of the transversal homoclinic orbit of $\mathcal{C}$ in the interior (resp. exterior) realm. Let A_1' and B_1' (resp. G_2' and H_2') be the first images of A_1 and B_1 (resp. G_2 and H_2) in $n_{1,1}$ (resp. $n_{2,2}$) sent by the forward flow outside $\mathcal{R}_1$ (resp. $\mathcal{R}_2$). The maps sending A_1, B_1, G_2, H_2 onto A_1', B_1', G_2', H_2' are diffeomorphisms. In a neighborhood of $p_{1,1}$ (resp. $p_{2,2}$) the qualitative picture of A_1' and B_1' (resp. G_2' and H_2') is shown in Figure 3.7.2 provided ϵ_1 and ϵ_2 are sufficiently small.

Similarly, let $p_{1,2} \in a_{1,2}^+$ and $p_{2,1} \in a_{2,1}^+$ be points of the transversal heteroclinic cycle of $\mathcal{C}$ in the Jupiter realm. Let A_2' and B_2' (resp. G_1' and H_1') be the first images of A_2 and B_2 (resp. G_1 and H_1) in $n_{1,2}$ (resp. $n_{2,1}$) sent by the flow outside $\mathcal{R}_1$ and $\mathcal{R}_2$. The maps sending A_2, B_2, G_1, H_1 onto A_2', B_2', G_1', H_1' are diffeomorphisms. In a neighborhood of $p_{1,2}$ (resp. $p_{2,1}$) the qualitative picture of A_2' and B_2' (resp. G_1' and H_1') is also shown in Figure 3.7.2.

Now let U_1 (resp. U_4) be the sets diffeomorphic to $(C_1 \cup D_1) \cap (A_1' \cup B_1')$ (resp. $(E_2 \cup F_2) \cap (G_2' \cup H_2')$) defined by following the flow backwards up to the first crossing with the surface $y = 0$. Similarly, let U_2 (resp. U_3) be the sets diffeomorphic to $(C_2 \cup D_2) \cap (G_1' \cup H_1')$ (resp. $(E_1 \cup F_1) \cap (A_2' \cup B_2')$) defined by following the flow backwards up to the first crossing with the surface $x = 1 - \mu$. See Figures 3.7.1 and 3.7.2. Since each of the sets U_i are topologically a square, we shall refer to them loosely as squares in the rest of this section.

Let $U = U_1 \cup U_2 \cup U_3 \cup U_4$. We define the Poincaré map

$$P : U \to U \tag{3.7.1}$$

in the following way: To each point $q \in U$ we assign the corresponding first intersection point with U of the orbit passing through q, if such an intersection exists. For simplicity of notation, we loosely refer to U_1 as

$(C_1 \cup D_1) \cap (A_1' \cup B_1')$ even though U_1 actually lies in the surface $y = 0$. Similar convention will be used for the other U_i's.

Now we consider the invariant set of points, Λ, which remain in U under all forward and backward iterations by P. Thus Λ is defined as

$$\Lambda = \cap_{n=-\infty}^{\infty} P^n(U). \tag{3.7.2}$$

This invariant set contains all the periodic solutions as well as the recurrent solutions near the chain and provides insight into the global dynamics in a neighborhood of the chain.

Horseshoe-type Map and Conley-Moser Conditions. We review a standard textbook example to introduce the next section. For a horseshoe-type map $h : Q \to Q$ of a square Q into itself, which satisfies the *Conley-Moser conditions*, the invariant set of all iterations

$$\Lambda_h = \cap_{n=-\infty}^{\infty} h^n(Q), \tag{3.7.3}$$

can be constructed and visualized in a standard way. The Conley-Moser conditions are the following.

- **Strip condition**: h maps "horizontal strips" H_0, H_1 to "vertical strips" V_0, V_1, (with horizontal boundaries to horizontal boundaries and vertical boundaries to vertical boundaries).

- **Hyperbolicity condition**: h has uniform contraction in horizontal direction and expansion in vertical direction.

The invariant set of first iterations

$$\Lambda_h^1 = h^{-1}(Q) \cap Q \cap h^1(Q), \tag{3.7.4}$$

has 4 squares, with addresses $(0;0), (1;0), (1;1), (0;1)$. Invariant set of second iterations has 16 squares contained in 4 squares of first stage. This process can be repeated ad infinitum due to the Conley-Moser conditions. What remains is invariant set of points Λ_h which are in 1-to-1 correspondence with set of bi-infinite sequences of 2 symbols $(\ldots, 0; 1, \ldots)$.

3.8 Horseshoe-like Dynamics

Compared with the standard textbook example above which studies the chaotic dynamics in a neighborhood of a transversal homoclinic point of a two-dimensional map $\bar{f}$, the Poincaré map P constructed in this chapter has a number of special properties.

Domain of the Poincaré Map P. Instead of studying the first return map $\bar{f}$ (induced by the flow f) on a (small) topological square Q, the domain U of the Poincaré map P consists of four squares $U_i, i = 1, 2, 3, 4$ which center around $p_{1,1}, p_{2,1}, p_{1,2}, p_{2,2}$, respectively, as shown in Figures 3.8.1 and 3.8.2.

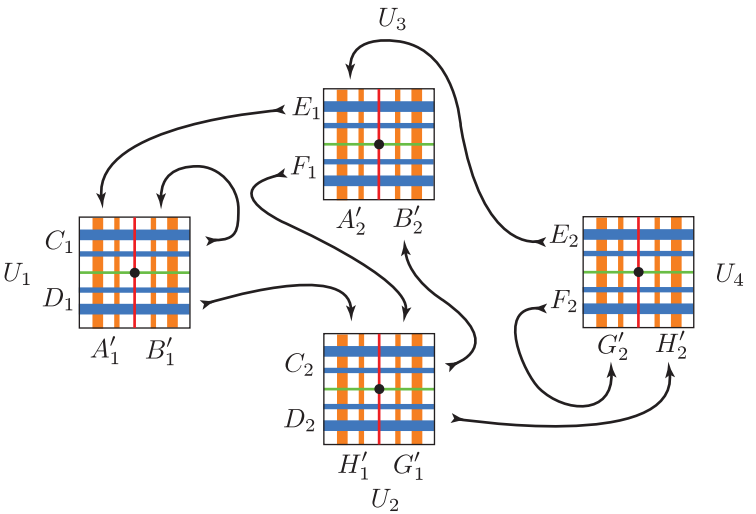


FIGURE 3.8.1. The families of horizontal strips and their images under P.

Moreover, the map P is not defined on points in U belonging to the invariant manifolds of the L_1 and L_2 Lyapunov orbits. Take U_1 as an example. On the curves $\Gamma^{u,\mathcal{S}}_{L_1,1}$ and $\Gamma^{s,\mathcal{S}}_{L_1,1}$ which are the first intersections of the unstable and stable invariant manifolds of the L_1 Lyapunov orbit with the surface $y = 0$ in the interior (Sun) realm, the Poincaré map is singular because any point on those curves will be carried by the flow asymptotically backward or forward towards the L_1 Lyapunov orbit. Hence, we have a kind of singular Poincaré map as it has been considered by Devaney [1981]. We return to this point at the end of §3.9.

Therefore, we must consider in fact four small (open) squares in U_1, namely:

$$(C_1 \cap A_1'), (C_1 \cap B_1'), (D_1 \cap A_1') \quad \text{and} \quad (D_1 \cap B_1').$$

Similar consideration is also needed for the other U_i's which add up to sixteen small squares in total, as shown in Figure 3.8.2.

Horizontal and Vertical Strips. For the standard textbook example, the first return map $\bar{f}$ (induced by the flow f) on the square Q qualitatively looks like a Smale horseshoe map. Conley and Moser found conditions for the map $\bar{f}$ to satisfy in order for it to have an invariant subset $\Lambda_{\bar{f}}$ of Q

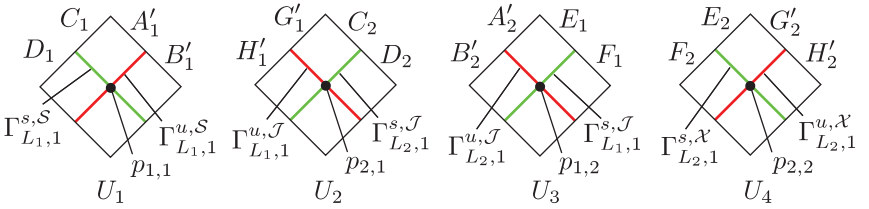


FIGURE 3.8.2. The domain $U = U_1 \cup U_2 \cup U_3 \cup U_4$ of the Poincaré map P.

on which it has chaotic dynamics. These conditions are a combination of geometrical and analytical conditions.

1. The geometrical part consists of generalizing the notion of horizontal and vertical rectangles to horizontal and vertical strips in Q by allowing the boundaries to be Lipschitz curves, as shown in Figure 3.8.3, rather than straight lines. With this generalization in hand one then requires "horizontal" strips to map to "vertical" strips with horizontal boundaries mapping to horizontal boundaries and vertical boundaries mapping to vertical boundaries.

2. The analytical part comes from requiring uniform contraction in the horizontal directions and expansion in the vertical direction.

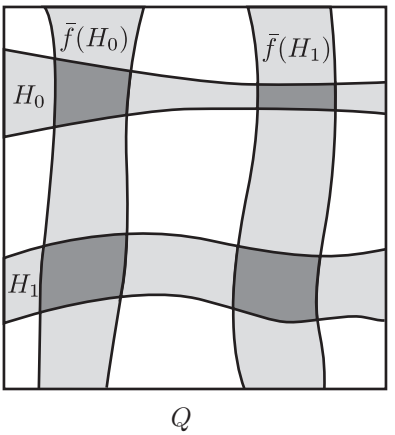


FIGURE 3.8.3. Generalization of the notion of horizontal and vertical rectangles for the Conley-Moser conditions.

For the Poincaré map P constructed in this chapter, the situation becomes more complicated in two ways. First, the number of strips in each family generated after one iteration is not two or even finite, but is instead infinite. Second, we need to use subshift to keep track of the image of each

family of strips. Here, we discuss first the issue of each family having an infinite number of strips.

First Iteration: 8 Families of Vertical Strips V_n^{ji}. Let us consider $U \cap P(U)$. For simplicity of exposition, take U_1 as an example and consider the small squares $(D_1 \cap A_1')$ and $(D_1 \cap B_1')$, shown on the left side of Figure 3.8.4.

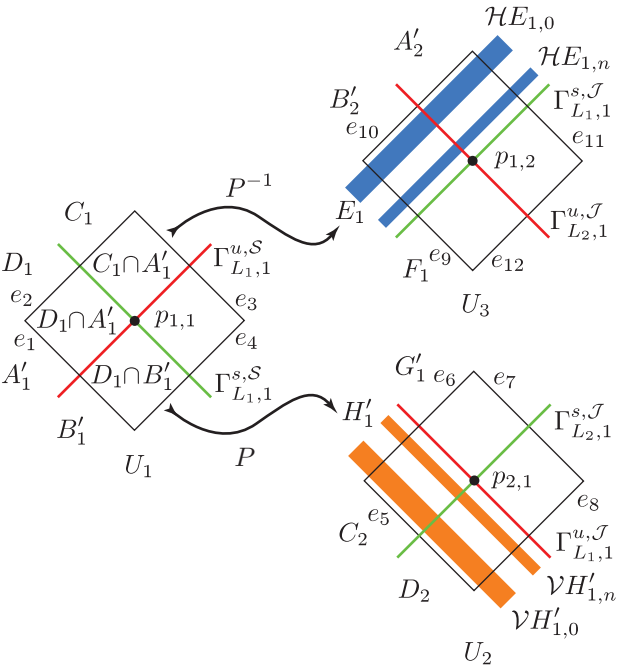


FIGURE 3.8.4. The topological squares and the images of some rectangles. We show schematically only two strips although there are an infinite number, getting increasingly slender as they approach the invariant manifold ($\Gamma^{u,\mathcal{J}}_{L_1,1}$ on U_2 and $\Gamma^{s,\mathcal{J}}_{L_1,1}$ on U_3).

Recall the observation in §2.8 on the spiraling of an abutting arc with an endpoint in the asymptotic set of a bounding sphere. The image of the squares $(D_1 \cap A_1')$ and $(D_1 \cap B_1')$ under P is a strip contained in H_1' of arbitrarily long length, cutting U_2 an infinite number of times and spiraling towards $\Gamma^{u,\mathcal{J}}_{L_1,1}$, becoming skinnier when approaching the limit. The intersection of this strip with U (in fact only with U_2) forms an infinite number of components. All but perhaps one of the components are limited by the sides e_6 and e_8, shown in Figure 3.8.4. We call each of the components of

$$P((D_1 \cap A_1') \cup (D_1 \cap B_1')) \cap U \subset H_1'$$

a ***vertical strip*** of H_1' (in U_2).

Now consider all the vertical strips in H_1' and denote these by $\mathcal{V}H_{1,0}', \mathcal{V}H_{1,1}', \ldots$, beginning with the strips nearest to e_5. We have on H_1' a family of vertical strips $\{\mathcal{V}H_{1,n}'\}$ bounded by the sides e_6 and e_8 (in U_2) and with the width of $\mathcal{V}H_{1,n}'$ tending to zero as n tends to infinity. We define

$$\mathcal{V}H_{1,\infty}' = \lim_{n\to\infty} \mathcal{V}H_{1,n}'.$$

Clearly, $\mathcal{V}H_{1,\infty}'$ is simply the vertical curve $\Gamma_{L_1,1}^{u,\mathcal{J}}$ which is on the Jupiter realm branch of the unstable invariant manifold of the L_1 Lyapunov orbit. Similar constructions can be carried out for the other small squares $(C_1 \cap A_1')$ and $(C_1 \cap B_1')$ of U_1 which yield a family of vertical strips in B_1'. In order to keep track of these families of vertical strips more effectively, we shall rename $\{\mathcal{V}B_{1,n}'\}$ and $\{\mathcal{V}H_{1,n}'\}$ as $\{V_n^{11}\}$ and $\{V_n^{21}\}$, respectively. Notice that for V_n^{ji}, the index ji indicates that the family is in the square U_j and it came from the square U_i. For simplicity of illustration, we have used rectangles to represent strips in Figure 3.8.4. Similar representations will be used throughout the rest of this section.

Similarly, we can look at the first iterate by P of the other U_i's and obtain families of vertical strips in

$$B_2'(\{V_n^{32}\}), H_2'(\{V_n^{42}\}), A_1'(\{V_n^{13}\}), G_1'(\{V_n^{23}\}), A_2'(\{V_n^{34}\}), G_2'(\{V_n^{44}\}).$$

Therefore, $U \cap P(U)$ is the disjoint union of eight families of pairwise disjoint vertical strips.

First Iteration: 8 Families of Horizontal Strips H_m^{ij}. An analogous study can be done for $U \cap P^{-1}(U)$. Consider the small squares $(D_1 \cap A_1')$ and $(C_1 \cap A_1')$ of U_1 in Figure 3.8.4. Then $P^{-1}((D_1 \cap A_1') \cup (C_1 \cap A_1'))$ is a strip contained in E_1 of arbitrarily long length, cutting U_3 an infinite number of times and spiraling towards $\Gamma_{L_1,1}^{s,\mathcal{J}}$, becoming thinner while approaching the limit. The intersection of this strip with U (in fact only with U_3) forms an infinite number of components. All but perhaps one of the components are limited by the sides e_9 and e_{11}. We call each of the components of

$$P^{-1}((D_1 \cap A_1') \cup (C_1 \cap A_1')) \cap U \subset E_1$$

a ***horizontal strip*** of E_1 (in U_3).

Now consider all the horizontal strips in E_1 and denote these by $\mathcal{H}E_{1,0}$, $\mathcal{H}E_{1,1}, \ldots$, beginning with the strip nearest to e_{10}. We have on E_1 a family of horizontal strips $\{\mathcal{H}E_{1,n}\}$ bounded by the sides e_9 and e_{11} (in U_3) and with the width of $\mathcal{H}E_{1,n}$ tending to zero as n tends to infinity. We define

$$\mathcal{H}E_{1,\infty} = \lim_{n\to\infty} \mathcal{H}E_{1,n}.$$

Clearly, $\mathcal{H}E_{1,\infty}$ is simply the horizontal curve $\Gamma_{L_1,1}^{s,\mathcal{J}}$ which is on the stable invariant manifolds of the L_1 Lyapunov orbit.

Similar constructions can be carried out for the other small squares $(C_1 \cap B_1')$ and $(D_1 \cap B_1')$ of U_1 which yield a family of horizontal strips in C_1. We shall again rename $\{\mathcal{H}C_{1,n}\}$ and $\{\mathcal{H}E_{1,n}\}$ as $\{H_n^{11}\}$ and $\{H_n^{31}\}$, respectively. Notice that for H_n^{ij}, the index ij indicates that the family is in the square U_i and it will go to the square U_j.

Similarly, we can look at the first iterate by P^{-1} of the other U_i's and obtain families of horizontal strips in

$$D_1(\{H_n^{12}\}), F_1(\{H_n^{32}\}), C_2(\{H_n^{23}\}), E_2(\{H_n^{43}\}), D_2(\{H_n^{24}\}), F_2(\{H_n^{44}\}).$$

Therefore, $U \cap P^{-1}(U)$ is the disjoint union of eight families of pairwise disjoint horizontal strips.

Meaning of the Label V_n^{ji}. We discuss briefly the meaning of the subscript n in the vertical strip V_n^{ji}. It can be used to keep track of the number of revolutions the projection of the associated orbits wind around L_1 or L_2. For example, the orbit which pierces the vertical strip V_{k+1}^{21} has wound one more time around L_1 than the orbit which pierces the vertical strip V_k^{21}. Moreover, given any ϵ_1 for the width of the strips D_1 and H_1', there is a minimum number of integer revolutions $r_{\min}$ around L_1 an orbit will make in going from D_1 (in U_1) to H_1' (in U_2). With this specific ϵ_1, the orbit which pierces V_n^{21} has wound around L_1 for $(n + r_{\min})$ times. In the rest of this chapter, we assume that we have adjusted the widths (the ϵ_j's) of all the other corresponding pairs of strips so that the minimum number of revolutions around L_1 or L_2 is the same for all the U_i's. With this adjustment, any orbit which pierces V_n^{ji} is now in U_j. It came from U_i and has wound around L_1 (if $u_i = 1, 3$) or L_2 (if $u_i = 2, 4$) for $(n + r_{\min})$ times.

The Generalized Conley-Moser Conditions. For the standard textbook example introduced earlier about the dynamics near a transversal homoclinic point, it is well known that if the first return map $\bar{f}$ (induced by f) on the square Q satisfies the following Conley-Moser conditions, then there exists an invariant set $\Lambda_{\bar{f}}$ of Q on which $\bar{f}$ has chaotic dynamics.

Condition 1: There exist a finite (or possibly infinite) number of horizontal and vertical strips $\mathcal{H}_i$ and $\mathcal{V}_i$ with i in an index set. The mapping $\bar{f}$ takes $\mathcal{H}_i$ homeomorphically onto $\mathcal{V}_i$, with horizontal boundaries mapped to horizontal boundaries and vertical boundaries mapped to vertical boundaries.

Condition 2: Suppose $\mathcal{V}$ is a vertical strip contained in $\bigcup_i \mathcal{V}_i$. Then $\bar{f}(\mathcal{V}) \cap \mathcal{V}_i = \bar{\mathcal{V}}_i$ is a vertical strip for every i. Moreover, $w(\bar{\mathcal{V}}_i) \leq \nu_v w(\mathcal{V})$ for some $0 < \nu_v < 1$ where $w(\mathcal{V})$ is the width of strip $\mathcal{V}$. Similarly, suppose $\mathcal{H}$ is a horizontal strip contained in $\bigcup_i \mathcal{H}_i$. Then $\bar{f}^{-1}(\mathcal{H}) \cap \mathcal{H}_i = \bar{\mathcal{H}}_i$ is a horizontal strip for every i. Moreover, $w(\bar{\mathcal{H}}_i) \leq \nu_h w(\mathcal{H})$ for some $0 < \nu_h < 1$.

In analogy with the conditions mentioned at the end of §3.7 for the horseshoe map, we call Condition 1 the ***strip condition***. Similarly, since Condition 2 requires a uniform contraction in the horizontal direction and expansion in the vertical direction, it can be called the ***hyperbolicity condition***.

For the Poincaré map P constructed in §3.7, the situation is more complex. Now we have four squares U_1 through U_4 together with eight families of pairwise disjoint horizontal strips and eight families of pairwise disjoint vertical strips. We state below the theorem that the Poincaré map P of the PCR3BP satisfies the ***generalized Conley-Moser conditions***.

3.8.1 Theorem. *The Poincaré map P satisfies the following generalized Conley-Moser conditions:*

Generalized Condition 1: *P maps horizontal strips to vertical strips, i.e.,*

$$\begin{array}{llll} P(H_n^{11}) = V_n^{11} & P(H_n^{12}) = V_n^{21} & P(H_n^{23}) = V_n^{32} & P(H_n^{24}) = V_n^{42} \\ P(H_n^{31}) = V_n^{13} & P(H_n^{32}) = V_n^{23} & P(H_n^{43}) = V_n^{34} & P(H_n^{44}) = V_n^{44} \end{array}$$

for all positive integers n, with horizontal boundaries mapping to horizontal boundaries and vertical boundaries mapping to vertical boundaries.

Generalized Condition 2: *Let V be a vertical strip contained in $\bigcup_i V_i^{13}$. Then*

$$V_n' = P(V) \cap V_n^{11} \quad \textit{and} \quad V_n'' = P(V) \cap V_n^{21}$$

are two vertical strips for every n. Moreover,

$$w(V_n') \le \nu_v w(V) \quad \textit{and} \quad w(V_n'') \le \nu_v w(V)$$

for some $0 < \nu_v < 1$, where $w(V)$ is the width of V. Similarly, let H be a horizontal strip contained in $\bigcup_i H_i^{11}$. Then

$$H_n' = P^{-1}(H) \cap H_n^{31} \quad \textit{and} \quad H_n'' = P^{-1}(H) \cap H_n^{11}$$

are two horizontal strips for every n. Moreover,

$$w(H_n') \le \nu_h w(H) \quad \textit{and} \quad w(H_n'') \le \nu_h w(H)$$

for some $0 < \nu_h < 1$. Similar assertions are true for the other families of vertical and horizontal strips.

The proof is in Koon, Lo, Marsden, and Ross [2000].
Recall that

$$\begin{array}{llll} \mathcal{H}C_{1,n} = H_n^{11} & \mathcal{H}D_{1,n} = H_n^{12} & \mathcal{H}E_{1,n} = H_n^{31} & \mathcal{H}F_{1,n} = H_n^{32} \\ \mathcal{H}C_{2,n} = H_n^{23} & \mathcal{H}D_{2,n} = H_n^{24} & \mathcal{H}E_{2,n} = H_n^{43} & \mathcal{H}F_{2,n} = H_n^{44} \\ \mathcal{V}A_{1,n}' = V_n^{13} & \mathcal{V}B_{1,n}' = V_n^{11} & \mathcal{V}G_{1,n}' = V_n^{23} & \mathcal{V}H_{1,n}' = V_n^{21} \\ \mathcal{V}A_{2,n}' = V_n^{34} & \mathcal{V}B_{2,n}' = V_n^{32} & \mathcal{V}G_{2,n}' = V_n^{44} & \mathcal{V}H_{2,n}' = V_n^{42}, \end{array}$$

where $\mathcal{H}C_{1,n}$ is the n-th horizontal strip of the horizontal rectangle C_1 and $\mathcal{V}A'_{1,n}$ is the n-th vertical strip of the vertical rectangle A'_1, etc. Moreover, the index ij of $\{H_n^{ij}\}$ indicates that the family is in the square U_i and it will go to the square U_j and the index ji of $\{V_n^{ji}\}$ indicates that the family is in the square U_j and it came from the square U_i, as illustrated in Figure 3.8.4.

We use this result to sketch the proof of the main theorem on the global orbit structure of the PCR3BP given in §3.9 and §3.10.

3.9 Symbolic Dynamics

In §3.7 and §3.8, we have constructed a Poincaré map P on U whose domain consists of four topological squares $U_i, i = 1, 2, 3, 4$, each of which is further subdivided into four smaller squares by two curves that lie on the invariant manifolds of the Lyapunov orbits. Moreover, P satisfies the generalized Conley-Moser conditions.

While we need to take stock of certain new features, the basic formalism developed by Smale, Conley and Moser still holds with a few modifications.

For the horseshoe map h which bends a square Q into a horseshoe and intersects it with the square, one has an infinite Cantor set of trapped points in the invariant set Λ_h, given earlier in (3.7.4),

$$\Lambda_h = \cap_{n=-\infty}^{\infty} h^n(Q),$$

which is the set of points in the square Q that remain in the square under all forward and backward iterations by h.

We can define an element of the invariant set by

$$p = \{q \in Q \mid h^i(q) \in H_{s_i}, i = 0, \pm 1, \pm 2, \ldots\},$$

where s_i denotes one of the elements in $\Sigma^2 = \{0, 1\}$ and H_0, H_1 are the two original horizontal rectangles in D. Moreover, an address which is a bi-infinite sequence of two symbols $\{0, 1\}$ (in Σ^2) can be attached to every point p in the invariant set Λ_h, which will not only describe its location, but also tell its whole history and future under iteration of the map. By this we mean that there is a map $\phi : \Lambda_h \to \Sigma^2$ defined by

$$\phi(p) = (\ldots, s_{-n}, \ldots, s_{-1}; s_0, s_1, \ldots, s_n, \ldots),$$

where $s_i = 0$ if $h^i(p) \in H_0$ and $s_i = 1$ if $h^i(p) \in H_1$.

One easy way to imagine the invariant set Λ_h is to draw the regions that remain trapped for one forward and one backward iteration in the square Q. This is the intersection of the thickest vertical and horizontal strips, so it is four squares lying in the corners of the original square. The

set trapped for two iterations forwards and two backwards is obtained by intersecting the thinner strips of these figures, yielding sixteen smaller squares contained in the four squares of the first stage, as shown in Figure 3.9.1. Notice the addresses that have been assigned to those squares. This process can be repeated ad infinitum. After infinitely many steps, what remains is a Cantor set of points which are in one-to-one correspondence with the set of bi-infinite sequences of two symbols $\{0, 1\}$.

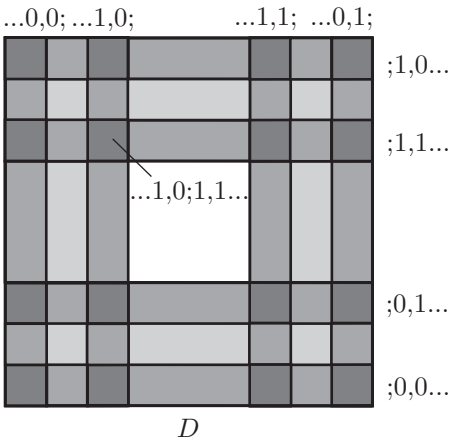


FIGURE 3.9.1. The invariant set Λ_h of the horseshoe map h.

For the Poincaré map P, we can use a similar technique to visualize the invariant set Λ and its associated set of bi-infinite sequences. Instead of one square Q, we have four squares $U_i, 1 = 1, 2, 3, 4$. After one forward and one backward iteration, instead of the intersections of two vertical rectangles and two horizontal rectangles, we have the intersections of eight families of vertical strips $\{V_n^{ji}\}$ and eight families of horizontal strips $\{H_n^{ij}\}$, with the indices ij corresponding to the nonzero entries of the transition matrix A. Using Figure 3.9.2 as a guide, recall from §3.7 that for $\{V_n^{ji}\}$, the index ji indicates that the family is in the square U_j and it came from the square U_i; for $\{H_n^{ij}\}$, the index ij indicates that the family is in the square U_i and it will go to the square U_j.

For simplicity of illustration, we draw Figure 3.9.2 schematically. Taking the family $\{H_n^{12}\}$ as an example, we draw two horizontal rectangles to represent the first and the n-th horizontal strips. This horizontal family is in the square U_1 and it will go to the square U_2. Similarly, for $\{V_m^{13}\}$, only the first and the m-th vertical rectangles are shown. This vertical family is in the square U_1 and it came from the square U_3. The same method has been used to illustrate all the other families of horizontal and vertical strips.

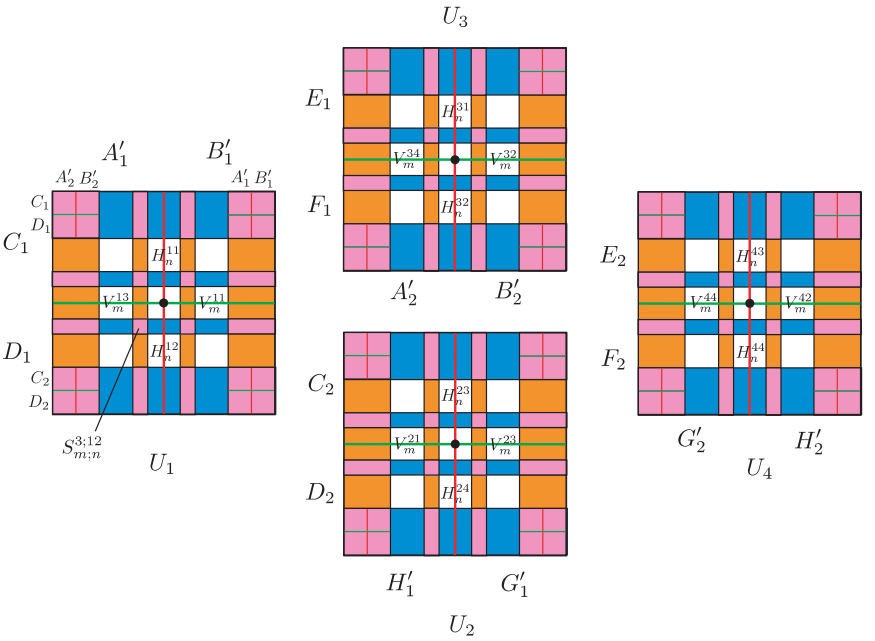


FIGURE 3.9.2. The invariant set Λ of the Poincaré map P.

As for assigning the addresses for points remaining in U, take the "square" $S_{m;n}^{3;12}$ as an example. Since $S_{m;n}^{3;12}$ is the intersection of the horizontal strip H_n^{12} and the vertical strip V_m^{13}, we can use $(\ldots, u_3, m; u_1, n, u_2, \ldots)$ to represent its location. As usual, the central block of this sequence also tells the history of the points in this "square" $(S_{m;n}^{3;12})$:

1. they are currently in U_1 and will go to U_2 and on their way their projection will wind around L_1 for $(n + r_{\min})$ revolutions where $r_{\min}$ is the minimum number of revolutions discussed earlier in §3.7;

2. they came from U_3 and their position space projection has wound around L_1 for $(m + r_{\min})$ revolutions.

Similar sequences can be assigned to the other "squares" which are the intersections of all the other horizontal and vertical strips.

Moreover, since the Poincaré map P satisfies the generalized Conley-Moser conditions, this process can be repeated ad infinitum as in the case of the horseshoe map. After an infinite number of steps, what remains in U is a Cantor set of points which are in one-to-one correspondence with the set of bi-infinite sequences

$$(\ldots, (u_{i_{-1}}, n_{-1}); (u_{i_0}, n_0), (u_{i_1}, n_1), (u_{i_2}, n_2), \ldots).$$

Hence, we have shown that the invariant set Λ for the Poincaré map P corresponds to a set of bi-infinite sequences with two families of symbols.

The first family is a subshift of finite type with four symbols $\{u_1, u_2, u_3, u_4\}$ (with a transition matrix A defined in §3.7). It is used to keep track of the history of the map P with respect to the four squares U_1, U_2, U_3, U_4.

The second family is a full shift of infinite type with symbols of non-negative integers. This set of integers is used to keep track of individual members of each vertical or horizontal family ($\{V_n^{ji}\}$ or $\{H_n^{ij}\}$). As mentioned at the end of §3.7, this set of integers also corresponds to the number of revolutions that the position space projection of an orbit winds around either L_1 and L_2.

Singular Poincaré Map. We discuss briefly the issue of the singular Poincaré map and how it relates to certain modifications of the space of symbol sequences Σ. Let $\Sigma = \{((u_{i_j}, n_j))\}$ be the set of bi-infinite sequences of elements of $S \times \mathbb{N}$ with a transition matrix A defined on S. Here, $S = \{u_1, u_2, u_3, u_4\}$ and $\mathbb{N}$ is the set of non-negative integers. As usual, a compactification $\bar{\Sigma}$ of Σ is obtained with the inclusion of sequences of the following types:

$$\begin{aligned}
\beta &= (\ldots; (u_{i_0}, n_0), \ldots, (u_{i_k}, \infty)) \\
\gamma &= (\infty, (u_{i_{-l}}, n_{-l}), \ldots; (u_{i_0}, n_0), \ldots) \\
\delta &= (\infty, (u_{i_{-l}}, n_{-l}), \ldots; (u_{i_0}, n_0), \ldots, (u_{i_k}, \infty)).
\end{aligned}$$

The elements of $\Sigma \subset \bar{\Sigma}$ will be called ***type*** α from now on. Moreover, the shift map σ on Σ defined by $\sigma((u_{i_j}, n_j)) = (u_{i_{j+1}}, n_{j+1})$ can be extended to a shift map $\bar{\sigma}$ in a natural way. The domain of $\bar{\sigma}$ is

$$D(\bar{\sigma}) = \{(u, n) \in \bar{\Sigma} \mid n_0 \neq \infty\}$$

and the range of $\bar{\sigma}$ is

$$R(\bar{\sigma}) = \{(u, n) \in \bar{\Sigma} \mid n_1 \neq \infty\}.$$

By studying Figure 3.9.2, it should be clear that H_∞^{12} (or H_∞^{11}) is simply the horizontal curve $\Gamma_{L_1,1}^{s,\mathcal{S}}$ which is on the interior (Sun) realm branch of the stable invariant manifold of the L_1 Lyapunov orbit and any point on this curve will be carried forward asymptotically towards the L_1 Lyapunov orbit. Hence, any element of type β corresponds to an orbit which tends to either the L_1 or L_2 Lyapunov orbit asymptotically after k iterations. Similarly, any element of type γ corresponds to an orbit which is carried by the flow asymptotically backward towards one of the Lyapunov orbits after l backward iterations. As for an element of type δ, we have either a homoclinic or a heteroclinic orbit.

3.10 Global Orbit Structure

Now we are ready to put together all the results in §3.8 and §3.9 and to state the main theorem of this chapter which provides a symbolic dynamics description of the global orbit structure of the PCR3BP near a chain of homoclinic orbits and a symmetric heteroclinic cycle. For simplicity of exposition, we have assumed in the past that the chain consists of $(1,1)$-homoclinic orbits in the interior and exterior realms and a symmetric $(1,1)$-heteroclinic cycle in the Jupiter realm. Now we consider the general situation. Let us suppose from now on that the chain $\mathcal{C}$ is made up of a symmetric (q_2,p_2)-heteroclinic cycle in the Jupiter realm together with two homoclinic orbits, one of which is a (q_1,p_1) orbit in the interior realm and the other is a (q_3,p_3) orbit in the exterior realm.

3.10.1 Theorem (Global Orbit Structure)**.** *Consider an element $(u,r) \in \bar{\Sigma}$ with $r_j \geq r_{\min}$ for all j. Then there are initial conditions, unique in a neighborhood of the given chain of two homoclinic orbits and one symmetric heteroclinic cycle (associated with $p_{1,1}, p_{2,2}, p_{1,2}, p_{2,1}$, respectively), such that the following statements are true.*

1. *For an element of type*

$$\alpha = (\ldots, (u_{i_{-1}}, r_{-1}); (u_{i_0}, r_0), (u_{i_1}, r_1), (u_{i_2}, r_2), \ldots),$$

the orbit corresponding to such conditions starts at U_{i_0} and goes to U_{i_1} if $(A)_{i_0 i_1} = 1$. This orbit passes through either the equilibrium region $\mathcal{R}_1$ or $\mathcal{R}_2$ depending on whether the initial index i_0 is $1,3$ or $2,4$. If $i_0 = 1,3$, the projection of the orbit winds around L_1 for r_0 revolutions inside the region $\mathcal{R}_1$ before leaving for U_{i_1}. Otherwise, it winds around L_2 for r_0 revolutions before leaving for U_{i_1}. After that, the same process begins with (u_{i_1}, r_1) in place of (u_{i_0}, r_0) and (u_{i_2}, r_2) in place of (u_{i_1}, r_1), etc. For negative time a similar behavior is described for $(u_{i_{-1}}, r_{-1}), (u_{i_0}, r_0)$, etc.

For this orbit, the number of revolutions that the comet winds around Jupiter or the Sun (in the interior or exterior realm) is a constant which depends on the realm and the given chain of homoclinic orbits and heteroclinic cycle. For the Jupiter realm, the number is $(q_2 + p_2 - 1)/2$. For the interior and exterior realms, the number is $q_1 + p_1 - 1$ and $q_3 + p_3 - 1$, respectively. Note that q_i and p_i are positive integers.

2. *For an element of type*

$$\beta = (\ldots; (u_{i_0}, r_0), \ldots, (u_{i_k}, \infty)),$$

the orbit tends asymptotically towards one of the Lyapunov orbits after k iterations. If $u_{i_k} = 1,3$, the orbit tends towards the L_1 orbit and stays in region $\mathcal{R}_1$. If $u_{i_k} = 2,4$, it tends towards the L_2 orbit and stays in region $\mathcal{R}_2$.

3. *For an element of type*

$$\gamma = (\infty, (u_{i_{-l}}, r_{-l}), \ldots; (u_{i_0}, r_0), \ldots),$$

the orbit tends asymptotically backward towards one of the Lyapunov orbits after l backward iterations. If $u_{i_{-l}} = 1, 2$, the orbit tends towards the L_1 orbit and stays in region $\mathcal{R}_1$. If $u_{i_{-l}} = 3, 4$, it tends towards the L_2 orbit and stays in region $\mathcal{R}_2$.

4. *For an element of type*

$$\delta = (\infty, (u_{i_{-l}}, r_{-l}), \ldots; (u_{i_0}, r_0), \ldots, (u_{i_k}, \infty)),$$

the orbit tends asymptotically towards the L_1 or L_2 Lyapunov orbit after k iteration, depending on whether $u_{i_k} = 1, 3$ or $2, 4$. It also tends asymptotically backward towards the L_1 or L_2 orbit after l iterations backwards, depending on whether $u_{i_l} = 1, 2$ or $3, 4$.

We provide a sketch of the proof here, which makes use of the major results in §3.8 and §3.9. The proof itself is in Koon, Lo, Marsden, and Ross [2000]. While we still need to fully establish the fact that the Poincaré map P does satisfy the generalized Conley-Moser conditions as mentioned at the end of §3.8, we refer the reader to the proof in Koon, Lo, Marsden, and Ross [2000] so that we can discuss the implications of this theorem.

Sketch of Proof. First construct a Poincaré map P whose domain U consists of four different squares $U_i, i = 1, 2, 3, 4$. Squares U_1 and U_4 are contained in the surface $y = 0$ and they center around (q_1, p_1) and (q_3, p_3)-transversal homoclinic points in the interior and the exterior realms, respectively. Squares U_2 and U_3 are contained in the surface $x = 1 - \mu$ and center around (q_2, p_2)-transversal heteroclinic points in the Jupiter realm which are symmetric with respect to each other.

Adjust the widths of all the corresponding pairs of the thin strips on the bounding spheres so that the minimum number of revolutions $r_{\min}$ around L_1 or L_2 is the same for all the U_i's. With this adjustment, any orbit which pierces V_m^{ji} is now in U_j. It came from U_i and has wound around L_1 (if $u_i = 1, 3$) or L_2 (if $u_i = 2, 4$) for $(m + r_{\min})$ times. A similar analysis holds for H_n^{ji}.

Assume that we have shown that the Poincaré map P satisfies the generalized Conley-Moser conditions. Then our discussion in §3.9 on symbolic dynamics shows that for any bi-infinite sequence of type α, $\alpha = (u, r)$, we can find initial conditions (u, n) in U such that the orbit with this initial condition has exactly the history of (u, r). Here, $r_j = n_j + r_{\min}$. Similar arguments also hold for bi-infinite sequences of other types.

Some Comments on the Implications of the Theorem. Type α orbits include "oscillating," "capture" and "non-transit" orbits. Oscillating orbits are orbits which cross from one realm to the other infinitely many

times, capture orbits are orbits which cross sometime but eventually stay in one realm, and non-transit orbits always stay in the same realm. Type β and type γ orbits are asymptotic orbits which wind to one of the Lyapunov orbits. Type δ orbits are homoclinic and heteroclinic orbits.

Similar to the standard textbook example, it is easy to verify that both the shift map $\bar{\sigma}$ and the Poincaré map P have the following properties:

1. a countable infinity of periodic orbits of all periods,
2. an uncountable infinity of nonperiodic orbits, and
3. a "dense orbit."

Moreover, both $\bar{\sigma}$ and P model the phenomenon that is called ***deterministic chaos*** in dynamical systems theory. Most notably, they exhibit the phenomenon of sensitive dependence on initial conditions, i.e., the distance between nearby initial conditions grows under some fixed number of iterates. This phenomenon corresponds to the "random" jumping of the comets between the interior, the Jupiter and the exterior realms.

4

Construction of Trajectories with Prescribed Itineraries

4.1 Introduction

Let us summarize the major results of the book to this point. We have been developing a framework for understanding the motion of a particle in the gravity field of two massive bodies, m_1 and m_2. In particular, we have considered the planar circular restricted three-body model. In this model, there is a constant of the motion, the energy, which divides the phase space of the particles motion into five cases (see Figure 2.4.3).

In the first four cases, depicted in Figure 2.4.2, the energy surface is naturally divided into three large realms of motion:

1. the m_1 realm, surrounding m_1;
2. the m_2 realm, surrounding m_2;
3. the exterior realm, which includes neither m_1 nor m_2, and is exterior to them both.

The energy cases are defined according to which realms of motion are connected. The connections appear as necks surrounding the location of libration points. The necks increase their width with increasing energy, corresponding to "easier" transport between realms with increasing energy. For example, in case 3, the particle has enough energy to move between all

three realms: the m_1, m_2, and exterior realms. The particle moves between realms via necks surrounding L_1 (connecting the m_1 and m_2 realms) and L_2 (connecting the m_2 and exterior realms). According to the terminology developed in Chapters 2 and 3, the neck regions surrounding L_1 and L_2 are denoted R_1 and R_2, respectively, as in Figure 4.1.1.

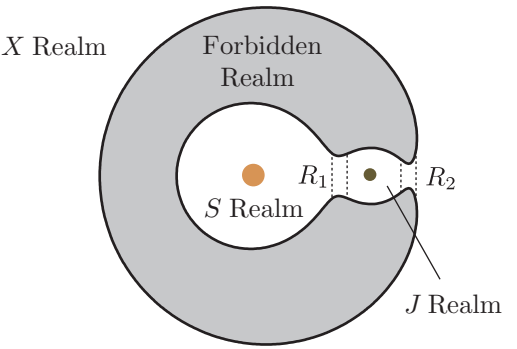


FIGURE 4.1.1. A schematic of the rotating frame for a particle in the gravitational field of the sun and Jupiter. Here, $m_1 = S$ and $m_2 = J$. Thus, the realm around the sun is the S realm, the realm around Jupiter is the J realm, and the realm not containing S or J is the X realm. One can construct orbits which connect the three realms using the stable and unstable manifold tubes associated to libration orbits in the necks around L_1 and L_2, the equilibrium regions R_1 and R_2, respectively (discussed in Chapters 2 and 3).

Tube Dynamics. Emanating from the periodic orbits in these necks are the stable and unstable manifolds with a $S^1 \times \mathbb{R}^1$ (cylindrical) geometry. The cylinders, or ***tubes***, have the physical property that all motion through the bottleneck in which the periodic orbit resides must occur through the interior of these surfaces. The tubes thus mediate the global transport of particles between large zones of the energy surface which are separated by the bottlenecks, i.e., the realms.

Since the tubes are global objects, we can, in theory, compute them out to arbitrarily long times and distances from the neck. Particles with initial conditions interior to a stable (respectively, unstable) manifold tube are guaranteed to move from one realm to another when evolved forward (resp., backward) in time. When one finds intersections between the regions interior to stable and unstable manifolds, one can pick any initial condition in the intersection region and integrate it both forward and backward. The resulting solution in the phase space corresponds to a desired particle path, i.e., a desired itinerary for the particle.

Outline of the Chapter. To find trajectories with prescribed itineraries, numerical methods are needed, namely for generating periodic orbits and

their invariant manifolds. §4.2 and §4.3 are devoted to laying the theoretical foundation for these purposes. §4.4 will outline an algorithm for using these methods to generate a trajectory with a prescribed itinerary in the planar CR3BP. An example computation is given in §4.5.

4.2 Differential Correction

In order to generate the invariant stable or unstable manifold of a periodic orbit, one first needs to compute the periodic orbit as accurately as possible. To do this, one can use method of ***differential correction***, which incorporates analytical approximations as the *first guess* in an iterative process aimed at producing initial conditions belonging to a periodic orbit.

Overview of Differential Correction. The idea of differential correction, or shooting (multiple or otherwise), is best described as a process of *targeting*. Suppose we want to find a periodic orbit. Given a reference trajectory $\bar{x}(t)$ going from $\bar{x}_0$ to $\bar{x}_1$ under the natural dynamics

$$\dot{x} = f(x), \quad x \in \mathbb{R}^n \tag{4.2.1}$$

e.g., the Hamiltonian equations of motion, we want to make slight adjustments, $\delta\bar{x}_0$, of the initial state $\bar{x}_0$ so that an adjusted trajectory will end up at a desired final state x_d, near $\bar{x}_1$. In order to find the correct adjustment, we need to know the sensitivity of changes in the final state $\delta\bar{x}_1$ to small changes in the initial state $\delta\bar{x}_0$. The linear approximation to this sensitivity is discussed below.

The Flow Map, State Transition Matrix, and Variational Equations. Let trajectories of the system (4.2.1) with $x(t_0) = x_0$ be denoted by $\phi(t, t_0)$. In other words, $\phi(t, t_0) : x(t_0) \mapsto x(t)$ denotes the ***flow map*** of the dynamical system (4.2.1), mapping particles from their initial location at time t_0 to their location at time t. For our purposes, we will denote the flow map as $\phi(t, t_0; x_0)$ or simply $\phi(t; x_0)$ so the dependence on the initial condition $x(t_0) = x_0$ is made clear. One can easily verify that the flow map satisfies the equations of motion (4.2.1),

$$\frac{d\phi(t; x_0)}{dt} = f(\phi(t; x_0)), \qquad \text{with} \quad \phi(t_0; x_0) = x_0.$$

A trajectory that starts slightly away from a reference trajectory $\bar{x}(t)$, i.e., starts from the perturbed initial vector $\bar{x}_0 + \delta\bar{x}_0$ at time t_0, will evolve with the displacement

$$\delta\bar{x}(t) = \phi(t; \bar{x}_0 + \delta\bar{x}_0) - \phi(t; \bar{x}_0) \tag{4.2.2}$$

with respect to the reference orbit $\bar{x}(t) = \phi(t; \bar{x}_0)$ as illustrated in Figure 4.2.1.

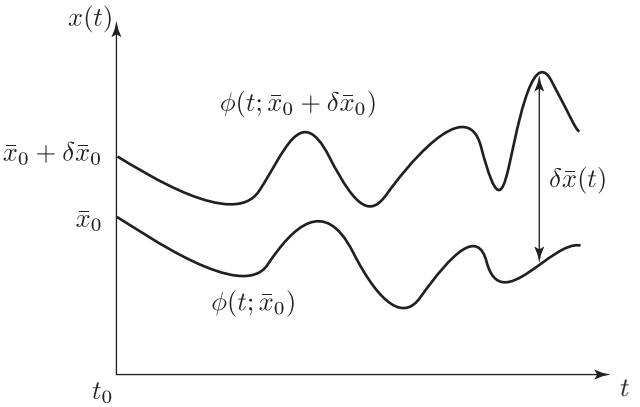


FIGURE 4.2.1. A trajectory $\phi(t; \bar{x}_0)$ and a neighboring trajectory, $\phi(t; \bar{x}_0 + \delta\bar{x}_0)$.

Measuring the displacement at time t_1 gives,

$$\delta\bar{x}(t_1) = \phi(t_1; \bar{x}_0 + \delta\bar{x}_0) - \phi(t_1; \bar{x}_0).$$

Expanding in a Taylor series yields

$$\delta\bar{x}(t_1) = \frac{\partial\phi(t_1; \bar{x}_0)}{\partial x_0}\delta\bar{x}_0 + \text{higher order terms.}$$

The matrix $\frac{\partial\phi(t_1;\bar{x}_0)}{\partial x_0}$ which satisfies the above relation to first-order is called the ***state transition matrix***. Usually abbreviated as $\Phi(t_1, t_0)$, this matrix gives the linear relationship between small initial and final displacements,

$$\delta\bar{x}(t_1) = \Phi(t_1, t_0)\delta\bar{x}_0. \tag{4.2.3}$$

and plays an important role in differential correction. Equation (4.2.3) is also the solution to the ***variational equations*** of (4.2.1),

$$\delta\dot{\bar{x}}(t) = Df(\bar{x}(t))\delta\bar{x}, \tag{4.2.4}$$

which are the linearized equations for the evolution of variations $\delta\bar{x}$. Here, $Df(\bar{x}(t))$ is the Jacobian matrix of the flow field f evaluated along the reference trajectory.

Differential Correction. Suppose we want to adjust an initial condition $\bar{x}_0$ in order to reach a desired target point x_d at time t_1. Without any adjustment,

$$\bar{x}(t_1) = \phi(t_1, t_0; \bar{x}_0) = \bar{x}_1 = x_d - \delta\bar{x}_1,$$

is slightly away from the desired point x_d (i.e., $|\bar{x}_1| > 0$ is small) and we want to correct it. Since,

$$\begin{aligned}\phi(t_1, t_0; \bar{x}_0 + \delta\bar{x}_0) &= \phi(t_1, t_0; \bar{x}_0) + \frac{\partial \phi(t_1, t_0; \bar{x}_0)}{\partial x_0}\delta\bar{x}_0 + O(|\delta\bar{x}_0|^2) \\ &= \phi(t_1, t_0; \bar{x}_0) + \Phi(t_1, t_0)\delta\bar{x}_0 + O(|\delta\bar{x}_0|^2) \\ &= \bar{x}_1 + \delta\bar{x}_1 + O(|\delta\bar{x}_0|^2) \\ &= x_d + O(|\delta\bar{x}_0|^2),\end{aligned}$$

we see that changing $\bar{x}_0$ by $\delta\bar{x}_0$ seems to make the correction to first-order. This is the basic idea behind differential correction: *making a small change at one end to target to a desired point at the other end*. By iteration, the process eventually produces convergence:

$$\phi(t_1, t_0; \bar{x}_0 + \Delta\bar{x}_0) = x_d + \varepsilon$$

where $\Delta\bar{x}_0$ is the accumulation of corrections $\delta\bar{x}_0$ which yields x_d within a desired tolerance $|\varepsilon| \ll 1$.

It is easy to see that a simple differential correction, or simple shooting, is just Newton's method of finding a root for the flow map. If we define:

$$F(x_0) = \phi(t_1, t_0; x_0) - x_d,$$

then targeting the desired final state is the same as finding the root of F. Newton's method calculates the root of F by iterating from an initial guess $\bar{x}_0$ using the iterative relation,

$$\delta\bar{x}_0 = -DF^{-1}(\bar{x}_0)F(\bar{x}_0),$$

which gives,

$$\delta\bar{x}_0 = \Phi(t_1, t_0)^{-1}\delta\bar{x}_1.$$

This is, of course, just equation (4.2.3).

Computation of the State Transition Matrix. To apply differential correction, one needs to compute the state transition matrix along a reference trajectory. Since $\phi(t; \bar{x}_0)$ satisfies the n-dimensional dynamical equations (4.2.1),

$$\frac{d\phi(t; \bar{x}_0)}{dt} = f(\phi(t; \bar{x}_0)),$$

with $\phi(t_0; \bar{x}_0) = \bar{x}_0$, differentiating this identity with respect to x_0 yields,

$$\frac{d}{dt}\frac{\partial \phi(t; \bar{x}_0)}{\partial x_0} = Df(\phi)\frac{\partial \phi(t; \bar{x}_0)}{\partial x_0},$$

and $\frac{\partial \phi(t_0; \bar{x}_0)}{\partial x_0} = I_n$, where I_n is the $n \times n$ identity matrix. Hence, the state transition matrix solves the following initial value problem,

$$\dot{\Phi}(t, t_0) = Df(\bar{x}(t))\Phi(t, t_0), \qquad \text{with} \quad \Phi(t_0, t_0) = I_n. \tag{4.2.5}$$

However, since $Df(\bar{x}(t))$ is time dependent, equations (4.2.5) cannot be solved analytically without an analytical solution for the reference trajectory. Since such a solution for the reference trajectory generally does not exist, equation (4.2.5) must be numerically integrated. This results in n^2 first-order scalar differential equations representing the elements of Φ. When combined with the n first-order scalar equations of motion (4.2.1), the result is a set of $n^2 + n$ differential equations requiring simultaneous numerical integration in order to compute the state transition matrix numerically.

4.3 Basic Theory on Invariant Manifolds of a Periodic Orbit

To study and compute the invariant manifolds of a periodic orbit, one needs two basic tools: the monodromy matrix and the Poincaré map. Below, we will use them to prove a number of important results that are related to the stability analysis of a periodic orbit. These results will provide the theoretical underpinning for the numerical computation of invariant manifolds. For more details, see Meyer and Hall [1992].

Monodromy Matrix. In what follows, we investigate the stability of one particular periodic solution $\bar{x}(t)$ with period T of the autonomous system (4.2.1) with $\bar{x}(0) = \bar{x}_0$.

Stability of $\bar{x}$ manifests itself in the way neighboring trajectories behave (refer to Figure 4.2.1). Referring to (4.2.2), a trajectory that starts from the perturbed initial vector $\bar{x}_0 + \delta\bar{x}_0$ will after one period T be displaced by

$$\delta\bar{x}(T) = \phi(T; \bar{x}_0 + \delta\bar{x}_0) - \phi(T; \bar{x}_0).$$

To first order, this displacement is given by

$$\delta\bar{x}(T) = \Phi(T)\delta\bar{x}_0,$$

where $\Phi(T) = \Phi(T; 0)$ is the state transition matrix after one period. This matrix, called the ***monodromy matrix*** M of the periodic orbit,

$$M \equiv \Phi(T) = \frac{\partial\phi(T; \bar{x}_0)}{\partial x_0} \tag{4.3.1}$$

determines whether initial perturbations $\delta\bar{x}_0$ from the periodic orbit decay or grow.

Basic Results on the Monodromy Matrix. Before studying the local stability of $\bar{x}(t)$, we will first state a few basic results (see, for example, Hartman [1964]) relating to the monodromy matrix:

- **Floquet theorem**: $\Phi(t) = P(t)e^{Rt}$ where $P(t)$ is T-periodic and R is a constant matrix.
- $\Phi(kT) = M^k$.
- M has $+1$ as an eigenvalue with eigenvector $f(\bar{x}_0)$ which is tangent to the periodic orbit at $\bar{x}_0$.

The Floquet theorem can be proved as follows: Since the Jacobian $Df(\bar{x})$ is periodic, it can be easily checked that for any matrix $\Phi(t)$ that solves

$$\dot{\Phi} = Df(\bar{x})\Phi,$$

$\Phi(t+T)$ is also a solution. Hence, there is a constant nonsingular matrix C such that

$$\Phi(t+T) = \Phi(t)C.$$

By using a logarithmic operator, C can be expressed in exponential form as

$$C = e^{RT} \tag{4.3.2}$$

where R is a constant matrix. Now let $P(t) = \Phi(t)e^{-Rt}$, and one can show that $P(t)$ is periodic:

$$P(t+T) = \Phi(t+T)e^{-R(t+T)} = \Phi(t)Ce^{-RT}e^{-Rt} = \Phi(t)e^{-Rt} = P(t).$$

Thus, the Floquet theorem is proved.

Notice that for the special Φ of equation (4.2.5), we have

$$\Phi(0+T) = \Phi(0)e^{RT} = e^{RT},$$

which shows $M = e^{RT}$. Clearly,

$$\Phi(kT) = e^{RkT} = (e^{RT})^k = M^k,$$

which proves the second result. Notice that this implies

$$\delta\bar{x}(kT) = M^k\delta\bar{x}_0 + \text{higher order terms}.$$

The eigenvalues λ of the monodromy matrix are called the *Floquet* or *characteristic multipliers*. Each number σ defined by $\lambda = e^{\sigma T}$ is called the *Floquet* or *characteristic exponent*. Both will be used in studying the local stability of the periodic solution.

As for the third result, it is obvious that if $\bar{x}(t)$ is a period T solution of equations (4.2.1), then the period T function $\dot{\bar{x}}(t)$ solves its variational equations, which we can write as

$$\dot{y} = Df(\bar{x})y,$$

for small displacements y away from the reference periodic solution $\bar{x}$. Since every solution of this linearized problem satisfies

$$y(t) = \Phi(t)y(0),$$

the relation

$$y(0) = y(T) = \Phi(T)y(0) = My(0)$$

shows that M has 1 as an eigenvalue. The eigenvector is

$$y(0) = \dot{\bar{x}}(0) = f(\bar{x}_0).$$

Poincaré Map and Stability Analysis of Periodic Orbits . Poincaré maps are useful for studying swirling flows, such as the flow near a periodic orbit. Let Σ be a co-dimension 1 surface of section. This hypersurface must be chosen such that all trajectories that cross Σ in a neighborhood of $q \in \Sigma$ meet two requirements: the trajectories intersect Σ transversely, and they cross Σ in the same direction.

For example, in the Sun-Jupiter-third body system, we might choose a section in the exterior realm as shown in Figure 4.3.1.

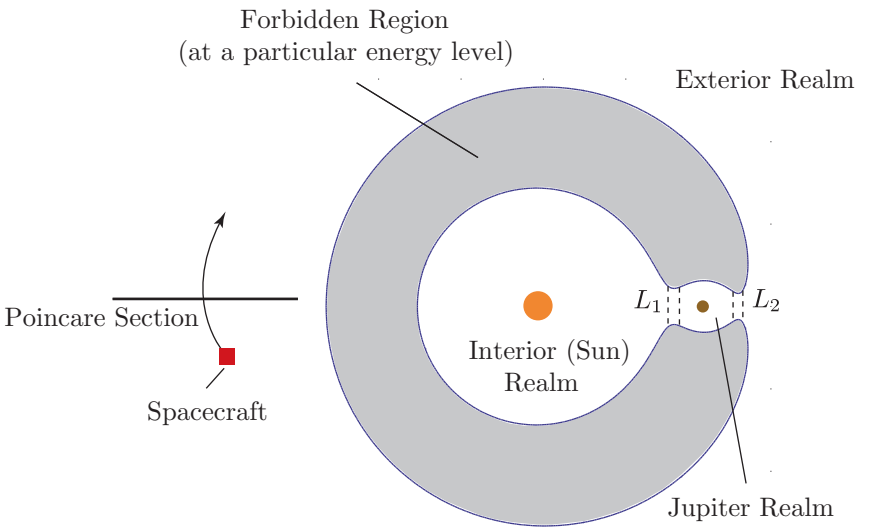


FIGURE 4.3.1. A Poincaré surface of section in the exterior realm of the Sun-Jupiter-third body system.

Such a procedure then produces a standard Poincaré map picture, as shown in Figure 4.3.2.

Let $T_\Sigma(q)$, with $q \in \Sigma$, be the time taken for a trajectory $\phi(t; q)$ to first return to Σ. The ***Poincaré map*** is defined by

$$P(q) \equiv \phi(T_\Sigma(q); q). \tag{4.3.3}$$

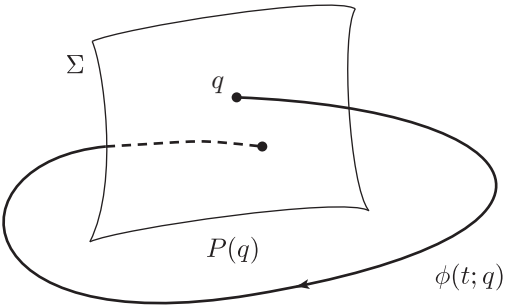


FIGURE 4.3.2. A Poincaré map produced by intersecting trajectories with a Poincaré section.

The periodic orbit is the specific trajectory that intersects Σ at $\bar{q}$ with $\bar{q} = \phi(T; \bar{q})$, i.e., $\bar{q}$ is a fixed point of P, $P(\bar{q}) = \bar{q}$.

Suppose we focus on the periodic orbit $\bar{x}(t)$, and in particular its stability. Its stability can be reduced to the behavior of the Poincaré map near its fixed point $\bar{q}$, as shown in Figure 4.3.3. Hence, the desired stability information of the periodic orbit is obtained by checking whether this fixed point $\bar{q}$ is repelling or attracting in the Poincaré map. Let $\frac{\partial P(\bar{q})}{\partial q}$ be the linearization of the discrete map P around the fixed point $\bar{q}$, given by

$$\frac{\partial P(\bar{q})}{\partial q} = \frac{\partial \phi(T; \bar{q})}{\partial q}, \tag{4.3.4}$$

and let $\lambda_1, \ldots, \lambda_{n-1}$ be the eigenvalues of this linearized map. If the moduli of all eigenvalues are smaller than 1, then $\bar{q}$ is stable; if the modulus of at least one eigenvalue is larger than 1, then $\bar{q}$ is unstable.

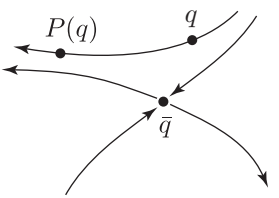


FIGURE 4.3.3. Poincaré map near a fixed point $\bar{q}$ which is of saddle or hyperbolic type, i.e., nearby point evolve along hyperbolic curves.

We now apply the stability result for a fixed-point to the periodic solution of equations (4.2.1). Recall that the monodromy matrix

$$M = \frac{\partial \phi(T; \bar{x}_0)}{\partial x_0}$$

has 1 as an eigenvalue with eigenvector $\dot{\bar{x}}_0$ tangent to the transversal curve $\bar{x}(t)$. Since this eigenvector is not in Σ, one can choose an appropriate basis

such that the remaining $(n-1)$ eigenvalues of M are those of $\frac{\partial P(\bar{q})}{\partial q}$. These eigenvalues are independent of the choice of Σ. As a result of this independence, the periodic orbit is stable if the remaining $(n-1)$ eigenvalues of M are smaller than 1 in modulus (the eigenvalues can be complex in general). The orbit $\bar{x}$ is unstable if M has an eigenvalue λ with $|\lambda| > 1$. The eigenvalue 1 corresponds to a perturbation along $\bar{x}(t)$ leading out of Σ; the other $(n-1)$ eigenvalues of M determine what happens to small perturbations around $\bar{q}$ within Σ.

Stability Result for Autonomous Hamiltonian Systems. For an autonomous Hamiltonian system, such as the CR3BP, which admits an energy integral E, we claim that the multiplier $+1$ has algebraic multiplicity at least 2. Moreover, the row vector $\frac{\partial E(\bar{x}_0)}{\partial x_0}$ is a left eigenvector of the monodromy matrix corresponding to the eigenvalue $+1$.

These results can be proven as follows: differentiating $E(\phi(t, x_0)) = E(x_0)$ with respect to x_0 and setting $x_0 = \bar{x}_0$ and $t = T$ yields

$$\frac{\partial E(\bar{x}_0)}{\partial x_0} \frac{\partial \phi(T, \bar{x}_0)}{\partial x_0} = \frac{\partial E(\bar{x}_0)}{\partial x_0}$$

which implies the second part of the result.

Choose coordinates such that $f(\bar{x}_0)$ is the column vector $(1, 0, \ldots, 0)^T$ and $\frac{\partial E(\bar{x}_0)}{\partial x_0}$ is the row vector $(0, 1, 0, \ldots, 0)$. Since $f(\bar{x}_0)$ is a right eigenvector and $\frac{\partial E(\bar{x}_0)}{\partial x_0}$ is a left eigenvector, the monodromy matrix M has the following form

$$M = \begin{pmatrix} 1 & * & * & * & \ldots & * \\ 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & * & * & * & \ldots & * \\ 0 & * & * & * & \ldots & * \\ \vdots & & & & & \vdots \\ 0 & * & * & * & \ldots & * \end{pmatrix}.$$

Expand by minors $p(\lambda) = \det(M - \lambda I)$. First expand along the first column to get $p(\lambda) = (1-\lambda)\det(M' - \lambda I)$, where M' is the $(n-1) \times (n-1)$ matrix obtained from M by deleting the first row and column. Expand $\det(M' - \lambda I)$ along the first row to get $p(\lambda) = (1-\lambda)^2 \det(M'' - \lambda I) = (1-\lambda)^2 q(\lambda)$, where M'' is the $(n-2) \times (n-2)$ matrix obtained from M by deleting the first two rows and columns. This computation proves the first part of the result.

Again there is a good geometric reason for the degeneracy implied by this result. The periodic solution lies in an $(n-1)$-dimensional level set of the integral, and typically in nearby level sets of the integral, there is a periodic orbit. So, periodic orbits are not isolated, as illustrated in Figure 4.3.4.

Consider the Poincaré map $P : N \to \Sigma$ where N is a neighborhood of $\bar{x}_0$ in Σ. Let u be the flow box coordinates so that $\bar{x}_0$ corresponds to $u = 0$;

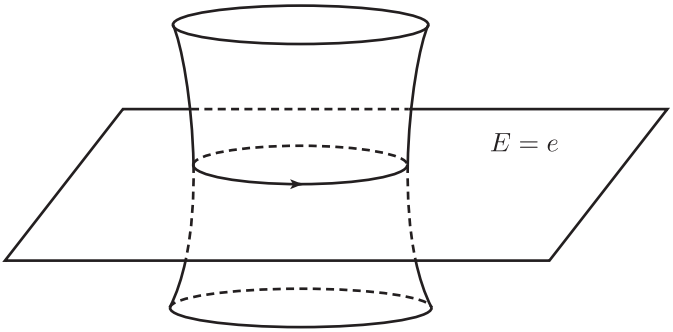


FIGURE 4.3.4. A family of periodic orbits on nearby integral surfaces.

equations (4.2.1) are

$$\dot{u}_1 = 1, \dot{u}_2 = 0, \ldots, \dot{u}_n = 0,$$

and $E(u) = u_2$. In these coordinates, we may take Σ to be $u_1 = 0$. Since u_2 is the integral in these coordinates, P maps the level set $u_2 =$ constant into itself; so we can ignore the u_2 component of P. Let $e = u_2$; let Σ_e be the intersection of Σ and the level set $E = e$; and let $y_1 = u_3, \ldots, y_{(n-2)} = u_n$ be coordinates in Σ_e (see Figure 4.3.5). Here, e is considered as a parameter

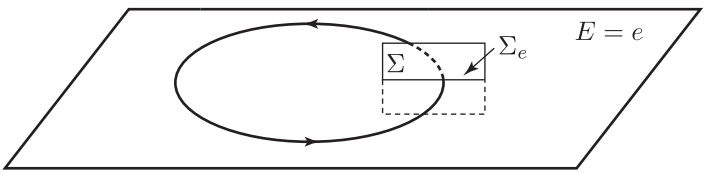


FIGURE 4.3.5. Poincaré map in an integral surface.

(the value of the integral). In these coordinates, the Poincaré map P is a function of y and the parameter e. So $P(e, y) = (e, Q(e, y))$, where for fixed e, $Q(e, \cdot)$ is a mapping of a neighborhood N_e of the origin in Σ_e into Σ_e. Q is called the Poincaré map in an integral surface. The eigenvalues of $\frac{\partial Q}{\partial y}(e, 0)$ are called the multipliers of the fixed point in the integral surface or the nontrivial multipliers. If the multipliers of the periodic solution of a system with nondegenerate integral are $1, 1, \lambda_3, \ldots, \lambda_n$, then the multipliers of the fixed point in the integral surface are $\lambda_3, \ldots, \lambda_n$.

Eigenvalues of the Monodromy Matrix for Hamiltonian Systems. Before actual calculation of invariant manifolds of a periodic orbit in the CR3BP, some important characteristics of the eigenvalues of the monodromy matrix of the CR3BP should be noted. From the discussion above, the monodromy matrix of a periodic orbit in the CR3BP must have at least two eigenvalues equal to one. Moreover, it can also be shown that

- The monodromy matrix M of an autonomous Hamiltonian system is symplectic.
- Hence, if λ is an eigenvalue of M, then $\lambda^{-1}, \bar{\lambda}, \bar{\lambda}^{-1}$ are also eigenvalues of M, with the same multiplicity

For an autonomous n degree of freedom Hamiltonian system with a Hamiltonian H, its Hamilton's equations can be written in a compact form as follows:

$$\dot{z} = J\nabla H(z), \qquad z = (q, p). \tag{4.3.5}$$

Here, J is a $2n \times 2n$ matrix defined by

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

where I_n denotes the $n \times n$ identity matrix.

Similar computation as in equations (4.2.4-4.2.5) show that the variational equation of (4.3.5) has the following form:

$$\dot{\Phi} = JS(\bar{z})\Phi, \qquad S(\bar{z}) = \frac{\partial^2 H}{\partial z^2}(\bar{z}) \qquad \text{with} \quad \Phi(0) = I_{2n}.$$

where $\Phi(t)$ is its fundamental matrix solution (the state transition matrix).

Now let $U(t) = \Phi(t)^T J\Phi(t)$. Since $\Phi(0) = I_{2n}$, it follows that $U(0) = J$. Moreover,

$$\begin{aligned} \dot{U}(t) &= \dot{\Phi}^T J\Phi + \Phi^T J\dot{\Phi} \\ &= \Phi^T S^T J^T J\Phi + \Phi^T JJS\Phi = \Phi^T(S^T - S)\Phi = 0 \end{aligned}$$

since S is symmetric. So, $U(t) \equiv J$. Therefore, $\Phi^T J\Phi = J$ and $\Phi(t)$ is symplectic. Since the monodromy matrix M is equal to $\Phi(T)$, we have the first proposition.

As for the second proposition, we begin with the following algebraic manipulations of the characteristic polynomial of a symplectic matrix A using simple properties of determinants.

$$\begin{aligned} p(\lambda) &= \det(A - \lambda I_{2n}) = \det(J(A - \lambda I_{2n})J^{-1}) \\ &= \det((A^{-1})^T - \lambda I_{2n}) = \det(A^{-1} - \lambda I_{2n}) \\ &= \det(A^{-1}(I_{2n} - \lambda A)) = \det(A^{-1})\det(I_{2n} - \lambda A) \\ &= \det(I_{2n} - \lambda A) = \det\left(-\lambda(A - \frac{1}{\lambda} I_{2n})\right) \\ &= \lambda^{2n}\det\left(A - \frac{1}{\lambda} I_{2n}\right) \\ &= \lambda^{2n} p\left(\frac{1}{\lambda}\right) \end{aligned}$$

Since det $A = 1$, it follows that 0 is not an eigenvalue of A. Therefore, it follows from our computation that if λ is an eigenvalue so is λ^{-1}. Moreover, since the coefficients of the characteristic polynomial are real (A is real), then if λ is an eigenvalue so is its complex conjugate $\bar{\lambda}$. For the issue of multiplicity, see Wiggins [2003].

4.4 Trajectories with Prescribed Itineraries

In this section, we go step by step through the construction of orbits with prescribed itineraries in the PCR3BP. We combine the geometric insight of the previous chapters with the methods and theory described in §4.2 and §4.3

For simplicity of exposition in the discussion which follows, consider the planar motion of a particle in the gravitational field of the sun and Jupiter ($\mu = 9.537 \times 10^{-4}$). We label the realm around the sun with an S, the realm around Jupiter with a J, and the exterior realm with an X, as in Figure 4.1.1. We will use the set of symbols $\{S, J, X\}$ to denote the location of the trajectory to construct finite itineraries of length k of the form $(A_1, A_2, \ldots, A_k)$, where $A_i \in \{S, J, X\}, i = 1, \ldots, k$. Using a conceptually simple procedure, trajectories with arbitrarily large itineraries can be constructed numerically.

A Trajectory with Itinerary (X, J, S). Suppose we want to find an initial condition corresponding to a particle which begins in the exterior realm and passes through the Jupiter realm to the sun realm. We transcribe this goal into a search for an initial condition with the itinerary (X, J, S). In principle, we could start with a large number of initial points in the four-dimensional phase space and save only those whose orbits correspond with this itinerary. But we can simplify the search tremendously by considering tube dynamics on an energy surface. Then our search becomes one of searching for an area on a two-dimensional Poincaré section for which all the points in that area correspond to an initial condition with this itinerary. We will use the following step by step procedure to find the itinerary region which corresponds to an orbit with itinerary (X, J, S). The itinerary region is a parcel, or lobe, on the Poincaré section

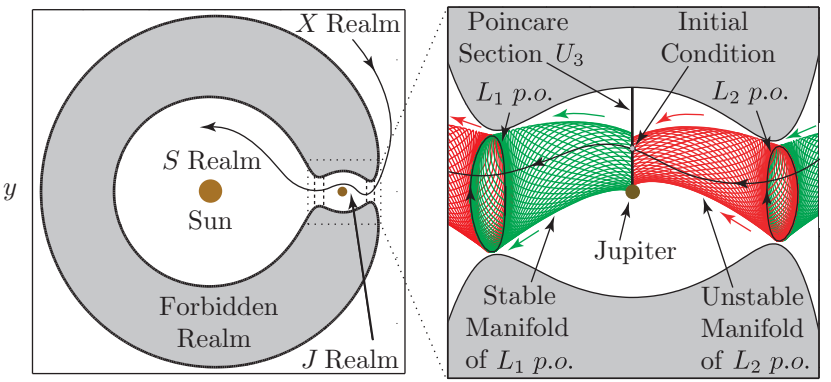
Algorithm for Finding an Itinerary Region. The reader may wish to reproduce the steps of this section in order to gain familiarity with the method.

Step 1. Select an Appropriate Energy. One first needs to set the energy to a value such that the itinerary one seeks exists. In our example, we want the particle to go between all three realms, X, J, and S, so we need to be in energy case 3 as described in §2.4. For the given μ, we compute

the case 3 energy interval,

$$(E_2, E_3) \approx (-1.519, -1.501).$$

For illustrative purposes, we will take a value $e \in (E_2, E_3)$ near the lower end of the interval. This corresponds to necks around L_1 and L_2 which are slightly open. The value we will use for the construction is $e = -1.515$. A schematic of the realms of possible motion for this energy is shown in Figure 4.4.1(a), and notice the labeling of the X, J, and S realms.



(a) (b)

FIGURE 4.4.1. (a) A schematic of the realms of possible motion for a case 3 energy. The X, J, and S realms are labeled. The trajectory shown as a heavy black line is a trajectory with an itinerary (X, J, S). (b) A close-up around the J-realm. The position space projection of the stable and unstable manifold tubes of the $L_i, i = 1, 2$ periodic orbits (p.o.) are shown. The J-branch of the L_1 stable (resp., L_2 unstable) tubes are labeled. We seek the intersection of the interior of these two tubes on the Poincaré section U_3. Taking an initial condition from this intersection and numerically integrating it forward and backward in time produces the trajectory with itinerary (X, J, S).

Step 2. Computing the Location of the Equilibrium Points. Consider the libration point L_i, standing for either L_1 or L_2. Compute the location of L_i, $(x_e,\ 0,\ 0,\ 0)$, using the procedure in §2.5. Consider the linearized equations of motion in a coordinate system centered on L_i, Eq. (2.6.4). The eigenvalues and eigenvectors for the linearized system are given by explicit formulas in §2.7. One can then compute, using the general solution (2.9.1) to (2.6.4), the initial conditions for a p.o. of x amplitude $A_x > 0$. In (2.9.1), let $t = \alpha_1 = \alpha_2 = 0$ and $\beta = -A_x/2$. When transformed back to the original coordinates, this yields an initial condition

$$\begin{aligned} \bar{x}_0 &= (x_e, 0, 0, 0) + 2\mathrm{Re}(\beta w_1), \\ &= (x_e - A_x, 0, 0, v_{y0}), \end{aligned} \tag{4.4.1}$$

where $v_{y0} = -A_x \nu\tau > 0$, $\nu = \sqrt{-\frac{1}{2}(\bar{\mu} - 2 - \sqrt{9\bar{\mu}^2 - 8\bar{\mu}})} > 0$, and $\tau = -(\nu^2 + 2\bar{\mu} + 1)/2\nu$, using relations from §2.7, where $\bar{\mu}$ is given in (2.6.2).

Step 3. Compute the L_1 and L_2 Periodic Orbits Using Differential Correction and Numerical Continuation. We consider one procedure which computes periodic orbits around L_i in a relatively straightforward fashion. This procedure begins with small "seed" periodic orbits obtained from the linearized equations of motion near L_i, and uses differential correction and numerical continuation to generate the desired periodic orbit (p.o.) corresponding to the chosen energy e.

The initial condition given in (4.4.1) will only yield a good approximation to a p.o. in the nonlinear equations (2.5.1) in the case of $A_x \ll 1$. But we want a p.o. of energy e, which may correspond to a large amplitude. We thus proceed as follows. Let $A_{x1} < A_{x2} \ll 1$ be two small amplitudes with corresponding initial conditions $\bar{x}_{0,g}^{(1)}$ and $\bar{x}_{0,g}^{(2)}$, respectively, where g denotes that this is an initial guess to a true periodic solution of (2.5.1).

We will use differential correction, described in §4.2, to produce initial conditions that lead to p.o.'s in the nonlinear equations which are accurate to some specified tolerance d. In other words, if $\bar{x}_{\rm po}(0) \equiv \bar{x}_0$ is an initial condition on a p.o., $x_{\rm po}(t)$, of period T, we want

$$|\bar{x}_{\rm po}(T) - \bar{x}_{\rm po}(0)| < d,$$

for a specified $d << 1$.

Differential correction uses an analytical approximation as the first guess in an iterative process which updates the initial conditions while keeping some values constant. In our case, we want to keep the x value constant and update the y velocity. The differential correction procedure is slightly modified here compared to §4.2 in order to incorporate sensitivity with respect to the final time.

Let trajectories of the differential equations $\dot{x} = f(x)$, e.g., (2.5.1), with $x(t_0) = x_0$ be denoted by the flow map $\phi(t, t_0; x_0)$, or $\phi(t; x_0)$ where the t_0 is understood. With $\bar{x}_0$ a first guess to an initial condition along a periodic orbit, a trajectory that starts from the perturbed initial vector $\bar{x}_0 + \delta\bar{x}_0$ evolves until $t + \delta t$ with the displacement

$$\delta\bar{x}(t + \delta t) = \phi(t + \delta t, t_0; \bar{x}_0 + \delta\bar{x}_0) - \phi(t, t_0; \bar{x}_0)$$

with respect to the first guess reference solution $\bar{x}(t)$.

Measuring the displacement at time $t_1 + \delta t_1$ gives

$$\delta\bar{x}(t_1 + \delta t_1) = \phi(t_1 + \delta t_1; t_0; \bar{x}_0 + \delta\bar{x}_0) - \phi(t_1, t_0; \bar{x}_0).$$

Expanding into Taylor series yields

$$\begin{aligned}\delta\bar{x}(t_1+\delta t_1) &= \frac{\partial\phi(t_1,t_0;\bar{x}_0)}{\partial x_0}\delta\bar{x}_0 + \frac{\partial\phi(t_1,t_0;\bar{x}_0)}{\partial t_1}\delta t_1 \\ &\quad +\text{higher order terms}, \\ &= \frac{\partial\phi(t_1,t_0;\bar{x}_0)}{\partial x_0}\delta\bar{x}_0 + \dot{\bar{x}}_1\delta t_1 + \text{h.o.t.},\end{aligned}$$

where the first part of the second term comes from $\frac{\partial\phi(t_1,t_0;\bar{x}_0)}{\partial t_1} = \frac{d\phi(t,t_0;\bar{x}_0)}{dt} = f(\phi(t,t_0;\bar{x}_0))$, evaluated at $t=t_1$. The matrix $\frac{\partial\phi(t_1,t_0;\bar{x}_0)}{\partial x_0}$ which satisfies the above relation to first order (when $\delta t_1 = 0$) is the state transition matrix $\Phi(t_1,t_0)$. The equation $\delta\bar{x}(t_1) = \Phi(t_1,t_0)\delta\bar{x}_0$ can be obtained numerically as the solution to the variational equations of the PCR3BP (2.5.1),

$$\delta\dot{\bar{x}}(t) = Df(\bar{x}(t))\delta\bar{x},$$

where the Jacobian matrix evaluated at $\bar{x}(t)$ is

$$Df(\bar{x}(t)) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\bar{U}_{xx} & -\bar{U}_{xy} & 0 & 2 \\ -\bar{U}_{yx} & -\bar{U}_{yy} & -2 & 0 \end{pmatrix}_{\bar{x}(t)},$$

and $\bar{U}_{ab}$ are the second partial derivatives of the effective potential (2.3.9).

Suppose we want to reach a desired endpoint, x_d, but

$$\bar{x}(t_1) = \phi(t_1,t_0;\bar{x}_0) = \bar{x}_1 = x_d - \delta\bar{x}_1,$$

is slightly off ($|\delta\bar{x}_1| > d$) and we need to correct it. Since

$$\begin{aligned}\phi(t_1,t_0;\bar{x}_0+\delta\bar{x}_0) &= \phi(t_1,t_0;\bar{x}_0) + \frac{\partial\phi(t_1,t_0;\bar{x}_0)}{\partial x_0}\delta\bar{x}_0 + \text{h.o.t.}, \\ &= \phi(t_1,t_0;\bar{x}_0) + \Phi(t_1,t_0)\delta\bar{x}_0 + \text{h.o.t.}, \\ &= \bar{x}_1 + \delta\bar{x}_1 + \text{h.o.t.}, \\ &= x_d + \text{h.o.t.},\end{aligned}$$

this implies that changing $\bar{x}_0$ by $\delta\bar{x}_0 = \Phi(t_1,t_0)^{-1}\delta\bar{x}_1$ will perform the correction to first order. By iteration, the process produces convergence:

$$|\phi(t_1,t_0;\bar{x}_0+\Delta\bar{x}_0) - x_d| < d,$$

where $\Delta\bar{x}_0$ is the accumulation of corrections $\delta\bar{x}_0$ which yields x_d within the desired tolerance d.

We seek periodic orbits which are symmetric w.r.t. the x-axis, noting that the symmetry

$$y \mapsto -y, t \mapsto -t$$

leaves the equations of motion (2.5.1) unchanged. The symmetry gives mirror image solutions $\bar{x}(-t)$ for each solution $\bar{x}(t)$, completing the other half of the periodic orbit.

From the earlier step, we choose an approximate initial condition (at $t_0 = 0$) which intersects the x-axis perpendicularly,

$$\bar{x}(0) = (x_0, 0, 0, v_{y0})^T.$$

Using a standard Runga-Kutta 7(8) integration package with an appropriate tolerance (say, 10^{-14}), we integrate this initial condition until the next x-axis crossing using the following procedure: (1) integrate until $y(t)$ changes sign; (2) then change the time step until, e.g., $|y(t)| < 10^{-11}$ which we refer to as the "crossing"; (3) at the crossing, $t_1 \equiv t$, $y_1 \equiv y(t_1)$.

This gives us $\bar{x}(t_1)$, so we can also compute $\Phi(t_1, 0)$. For a p.o., the desired final state has the form

$$\bar{x}(t_1) = (x_1, 0, 0, v_{y1})^T,$$

where $t_1 = T/2$, a half-period of the p.o. The actual value for v_{x1} as a result from numerical integration may not be 0. For our purposes, we want $|v_{x1}| < d$, where, e.g., $d = 10^{-8}$. The state transition matrix after one half-cycle, $\Phi(t_1, 0)$, can be used to adjust the initial values to obtain a p.o. as

$$\delta\bar{x}_1 \approx \Phi(t_1, 0)\delta\bar{x}_0 + \dot{\bar{x}}_1 \delta t_1.$$

Suppose $|v_{x1}| > d$ and we hold x_0 fixed. The correction to v_{y0} can be calculated from

$$\begin{aligned} \delta v_{x1} &= \Phi_{34}\delta v_{y0} + \dot{v}_{x1}\delta t_1 + \text{h.o.t.}, \\ 0 = \delta y_1 &= \Phi_{24}\delta v_{y0} + v_{y1}\delta t_1 + \text{h.o.t.}, \end{aligned}$$

where Φ_{ij} is an element of the matrix $\Phi(t_1, 0)$ and $\dot{v}_{x1}$ comes from the equations of motion (2.5.1) evaluated at the crossing. Here, $\delta v_{x1} = v_{x1}$ since we want $v_{x1} = 0$. Hence the y-velocity should be adjusted by

$$\delta v_{y0} \approx \left(\Phi_{34} - \frac{v_{x1}}{v_{y1}}\Phi_{24}\right)^{-1} v_{x1},$$

to cancel out v_{x1} if we let

$$v_{y0} \mapsto v_{y0} - \delta v_{y0}.$$

This process converges to $|v_{x1}| < d$ within a few iterations typically.

The above procedure yields an accurate initial condition for a p.o. from a single initial guess. If our initial guess came from the linear approximation near the equilibrium point (from §2.9), it has been observed numerically that we can only use this procedure for small amplitude p.o.'s around

L_i, e.g., amplitudes of order 10^{-4} for $\mu = 9.537 \times 10^{-4}$ (the Sun-Jupiter system). But if we want an orbit of arbitrarily large amplitude (which is in one-to-one correspondence with the energy e), we need to use ***numerical continuation*** to generate a family of orbits which reaches the appropriate energy e.

We proceed as follows. Suppose we find two small nearby p.o. initial conditions, $\bar{x}_0^{(1)}, \bar{x}_0^{(2)}$, correct to within the tolerance d, using the differential correction procedure described above. We can generate a family of p.o.'s with increasing amplitude around L_i in the following way. Let

$$\begin{aligned}\Delta &= \bar{x}_0^{(2)} - \bar{x}_0^{(1)}, \\ &= (\Delta x_0, 0, 0, \Delta v_{y0})^T.\end{aligned}$$

Extrapolate to an initial guess for $\bar{x}_0^{(3)}$ via

$$\begin{aligned}\bar{x}_{0,g}^{(3)} &= \bar{x}_0^{(2)} + \Delta, \\ &= \left((x_0^{(2)} + \Delta x_0), 0, 0, (v_{y0}^{(2)} + \Delta v_{y0})\right)^T \\ &= \left(x_0^{(3)}, 0, 0, v_{y0}^{(3)}\right)^T.\end{aligned}$$

Keeping $x_0^{(3)}$ fixed, we can use differential correction to compute an accurate solution $\bar{x}_0^{(3)}$ from the initial guess $\bar{x}_{0,g}^{(3)}$ and repeat the process until we have a family of solutions. We can keep track of the energy of each p.o. and when we have two solutions, $\bar{x}_0^{(k)}, \bar{x}_0^{(k+1)}$, whose energies bracket the desired energy e, we can refine our continuation until we find a p.o. of energy e to within a desired amount. Let X_0 denote the initial condition on the desired p.o.

Step 4. Computation of Invariant Manifolds. First we find the local approximations to the unstable and stable manifolds of the p.o. from the eigenvectors of the monodromy matrix. Next, the local, linear approximation of, for example, the unstable manifold, in the form of a state vector, is integrated in the nonlinear equations of motion (2.5.1) to produce the approximation of the stable and unstable manifolds, a procedure known as *globalization of the manifolds*, which we outline below.

The state transition matrix $\Phi(t)$ along the p.o. can be obtained numerically by integrating the variational equations from time 0 to T (see §4.2). Once the monodromy matrix $M = \Phi(T)$ is obtained, its eigenvalues can be computed numerically. For planar Lyapunov orbits in the PCR3BP, the discussion of §4.3 tells us that the four eigenvalues of M include one real pair and one pair equal to unity,

$$\lambda_1 > 1, \qquad \lambda_2 = \frac{1}{\lambda_1}, \qquad \lambda_3 = \lambda_4 = 1.$$

The eigenvector associated with eigenvalue λ_1 is in the unstable direction; the eigenvector associated with eigenvalue λ_2 is in the stable direction. Let $Y^s(X_0)$ denote the normalized (to 1) stable eigenvector, and $Y^u(X_0)$ denote the normalized unstable eigenvector. We now use these to compute the approximate manifolds as follows: Let

$$X^s(X_0) = X_0 + \epsilon Y^s(X_0) \tag{4.4.2}$$

be the initial guess for the stable manifold at X_0 along the p.o., as illustrated in Figure 4.4.2, and let

$$X^u(X_0) = X_0 + \epsilon Y^u(X_0)$$

be the initial guess for the unstable manifold at X_0. Here ϵ is a small

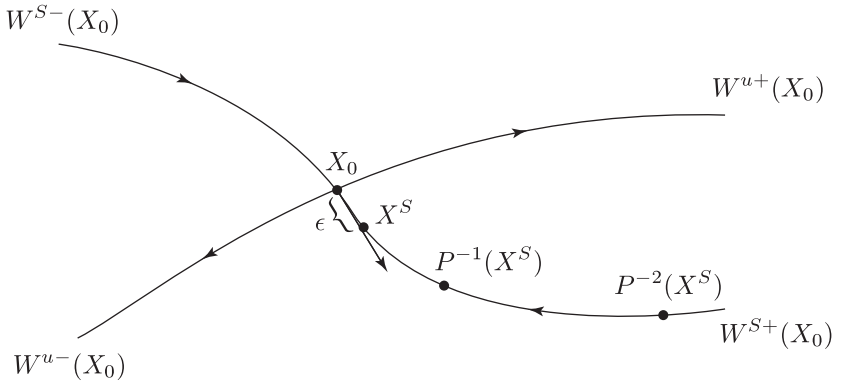


FIGURE 4.4.2. A simple way to compute an approximation of the two branches of the unstable ($W^{u\pm}$) or stable ($W^{s\pm}$) manifolds of a periodic orbit, which appear as manifolds to a fixed point in a Poincaré map P.

displacement from X_0. The magnitude of ϵ should be small enough to be within the validity of the linear estimate, yet not so small that the time of flight becomes too large due to the asymptotic nature of the stable and unstable manifolds. Gómez, Jorba, Masdemont, and Simó [1991a] suggest values of $\epsilon > 0$ corresponding to nondimensional position displacements of magnitude around 10^{-6}. By numerically integrating the unstable vector forwards in time, using both ϵ and $-\epsilon$, we generate trajectories shadowing the two branches, W^{u+} and W^{u-}, of the unstable manifold of the periodic orbit. Similarly, by integrating the stable vector backwards, we generate a trajectory shadowing the stable manifold, $W^{s\pm}$. For the manifold at $X(t)$, one can simply use the state transition matrix to transport the eigenvectors from X_0 to $X(t)$:

$$Y^s(X(t)) = \Phi(t, 0)Y^s(X_0). \tag{4.4.3}$$

Since the state transition matrix does not preserve the norm, the resulting vector must be renormalized.

Step 5. Take a Poincaré Surface of Section of the Globalized Stable and Unstable Manifolds. In Chapter 3, we defined the four Poincaré surfaces of section, $U_i, i = 1, \ldots, 4$, which were locally defined around heteroclinic points, with a map P linking them.

In order to link the present numerical construction with the earlier theoretical framework and terminology, we adopt the following convention. The U_1 and U_4 (Poincaré) sections will be defined by the following two-dimensional surfaces:

$$U_1 = \{(x, \dot{x}) \mid y = 0,\ x < 0,\ \dot{y}(x, \dot{x}; e) < 0\}, \text{ in the } S \text{ realm};$$

$$U_4 = \{(x, \dot{x}) \mid y = 0,\ x < -1,\ \dot{y}(x, \dot{x}; e) > 0\}, \text{ in the } X \text{ realm},$$

where $\dot{y}(x, \dot{x}; e)$ denotes that $\dot{y}$ is obtained from the energy equation (2.3.13). The U_2 and U_3 sections will be defined by the following:

$$U_2 = \{(y, \dot{y}) \mid x = 1 - \mu,\ y < 0,\ \dot{x}(y, \dot{y}; e) > 0\}, \text{ in the lower half } J \text{ realm};$$

$$U_3 = \{(y, \dot{y}) \mid x = 1 - \mu,\ y > 0,\ \dot{x}(y, \dot{y}; e) < 0\}, \text{ in the upper half } J \text{ realm}.$$

Figure 4.4.3 depicts the locations of the Poincaré sections in the rotating frame.

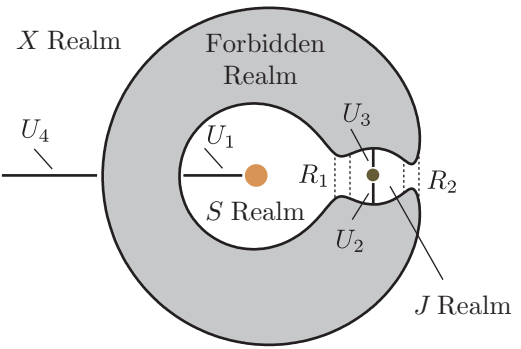


FIGURE 4.4.3. The location of the four Poincaré sections U_1, U_2, U_3, and U_4, with respect to the S, J, and X realms and the neck regions, R_1 and R_2, connecting them.

The U_i are at strategically placed locations, allowing us to get cross sections of the flow within the three-dimensional energy surface $\mathcal{M}(\mu, e)$. To pick the appropriate U_i on which to find an (X, J, S) lobe, we reason as follows. From our discussions regarding the L_1 and L_2 p.o. stable and unstable manifold tubes in Chapters 2 and 3, we know that, in a frame parallel to the rotating frame but centered on the point L_i, the two unstable manifold tube branches are locally heading in the second and fourth quadrants. Similarly, the stable manifold tube branches are locally coming from the first and third quadrants. (The reader may wish to review Figure

2.9.1 and §2.9.) Thus, the location of, for example, U_3, is chosen to intersect both the stable manifold of the L_1 p.o. and unstable manifold of the L_2 p.o., as shown in Figure 4.4.1.

Aside: Why Does this Method Work?

Recall from Chapters 2 and 3 the McGehee representation of the equilibrium region $\mathcal{R}$, which is in between two realms, e.g., the S and J realms. Emanating from the unstable p.o. are four cylinders of asymptotic orbits which form pieces of the stable and unstable manifold tubes of the p.o. They intersect the bounding spheres at asymptotic circles, separating spherical caps, which contain transit orbits, from spherical zones, which contain non-transit orbits. In order for an initial condition $s_0 \in \mathcal{R}$ to transit from one realm to another, it must be inside the tubes.

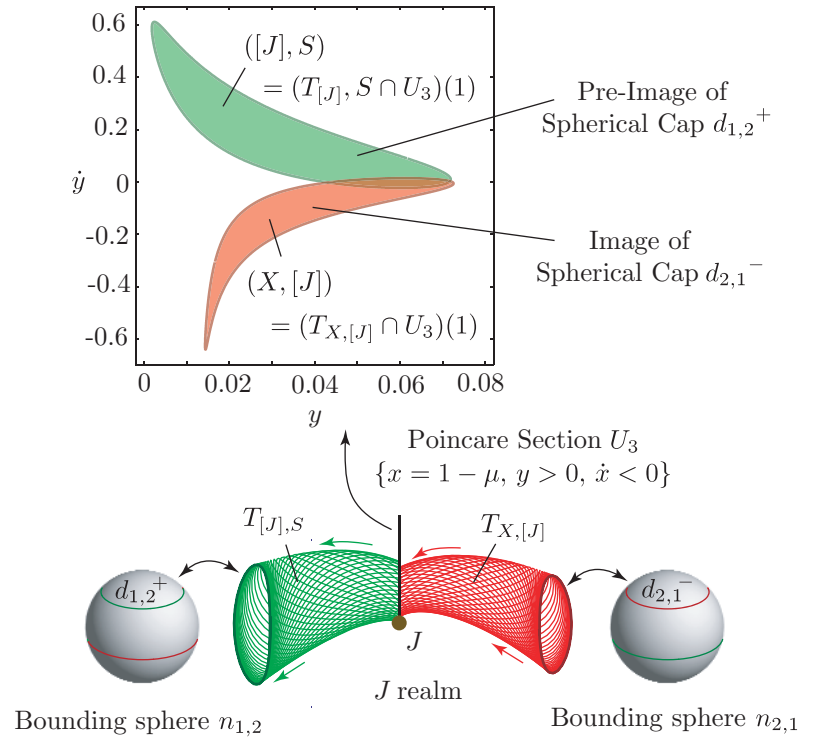


FIGURE 4.4.4. We seek transit orbits from the exterior to interior realm by looking at the intersections of images and pre-images of the "caps of transit orbits," introduced for the equilibrium regions in Chapters 2 and 3. See the text for an explanation.

For a fixed energy in case 3, consider the spherical caps of transit orbits. These are building blocks from which we construct orbits of prescribed

itineraries. Consider their images and pre-images on a suitable Poincaré section, e.g., the Poincaré section U_3 between L_1 and L_2. Spherical caps have the geometry of a disk, so we expect the images and pre-images will also appear as disk, or distorted disks. In Figure 4.4.4, the image of the cap on the left bounding sphere of the L_2 equilibrium region $\mathcal{R}_2$ is shown, containing trajectories leaving $\mathcal{R}_2$. On this same figure, we show the pre-image of the cap on the right bounding sphere of $\mathcal{R}_1$ containing orbits entering $\mathcal{R}_1$.

The intersection of the unstable manifold tube of the L_2 p.o. with U_3 forms the boundary of the image of the cap containing transit orbits leaving $\mathcal{R}_2$. All of these orbits came from the X realm and are now in the J realm, so we label this region $(X, [J])$.

Similarly, the intersection of the stable manifold tube of the L_1 p.o. with U_3 forms the boundary of the pre-image of the cap of transit orbits entering $\mathcal{R}_1$. All of these orbits are now in the J realm and are headed for the S realm, so we label this region $([J], S)$.

Note that the regions $(X, [J])$ and $([J], S)$ intersect.

Step 6. Consider Tube Dynamics to Compute the Desired Itinerary Region. From our discussion in Chapter 3, we know that the stable and unstable manifold tubes of the L_1 and L_2 p.o.'s bound regions in the energy surface exactly corresponding to motion between realms.

Key to our construction is the connectivity of the stable and unstable manifolds of the L_1 and L_2 p.o.'s. Consider Figure 4.4.5, where we show their projection onto position space which appear as strips of variable width. We show the tube projections up to their first intersection with the U_i.

For convenience in the discussion which follows, we introduce a new labeling convention. The set $T_{[A],B}$ is the *solid tube* of trajectories which are currently in the A realm and heading toward the B realm. The boundary of $T_{[A],B}$ is the stable manifold of the p.o. lying in the neck between the A and B realms. Similarly, the set $T_{A,[B]}$ is the solid tube of trajectories which came from the A realm and are currently in the B realm and its boundary is the unstable manifold of the p.o. lying in the neck between the A and B realms.

Consider the J realm. Suppose the initial condition for the trajectory we want to construct with itinerary (X, J, S) is in this realm. Then the itinerary region is $(X, [J], S)$. All particles with initial conditions in this piece of the Poincaré section are such that when numerically integrated backward in time they transit to the X realm, and when integrated forward in time they transit to the S realm, like the trajectory shown in Figure

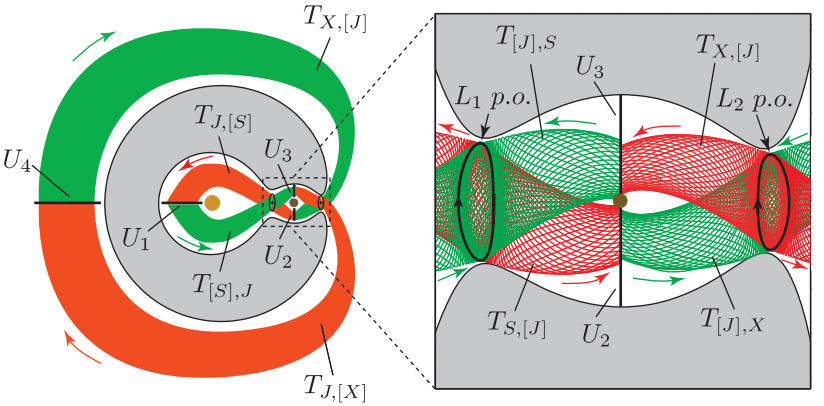


FIGURE 4.4.5. Position space projection of the L_1 and L_2 periodic orbit stable and unstable manifold tubes (schematic). The tubes are labeled according to the behavior of trajectories inside the boundaries defined by the stable and unstable manifolds. For example, $T_{[X],J}$ contains trajectories which are currently in the X realm and heading toward the J realm. Note the symmetry about the sun-Jupiter line. The location of the Poincaré surfaces of sections U_i are also shown. Magnification of the J realm is shown at right.

4.4.1. During the backward integration segment, the trajectory was within the J-branch of the L_2 p.o. unstable tube, labeled $T_{X,[J]}$ in Figure 4.4.5. Similarly, during the forward integration segment, the trajectory will be within the J-branch of the L_1 p.o. stable tube, labeled $T_{[J],S}$ in Figure 4.4.5. These two tubes are known from numerical experiments to intersect on the U_3 Poincaré section.

$T_{X,[J]}$ may contain pieces which wind around Jupiter several times. Therefore, $T_{X,[J]}$ will intersect U_3 several times. We denote the nth intersection of $T_{X,[J]}$ with U_3 by $\left(T_{X,[J]} \bigcap U_3\right)(n)$. For the present, we will restrict ourselves to $n = 1$.

The set $\left(T_{X,[J]} \bigcap U_3\right)(1)$ is a lobe of particles which came from the X realm and are now in J realm. Let us denote it by $I_{X,[J]}$, or simply $(X,[J])$, as in Figure 4.4.6. In Figure 4.4.6, we also plot $([J],S) = \left(T_{[J],S} \bigcap U_3\right)(1)$. We denote the intersection $(X,[J]) \bigcap ([J],S)$ by $(X,[J],S)$. This region contains initial conditions for orbits with itinerary (X,J,S), like the one shown in Figure 4.4.1.

Step 7. Numerically Integrate an Initial Condition in the Appropriate Itinerary Region. Once we have isolated the desired itinerary region, the last step is forward and backward numerical integration of any initial condition within this region. Continuing the example, suppose we have obtained the set $(X,[J],S)$, a subset of the y-$\dot{y}$ plane in U_3. We desire an initial condition $s_0 = (x_0\ y_0\ \dot{x}_0\ \dot{y}_0)^T$.

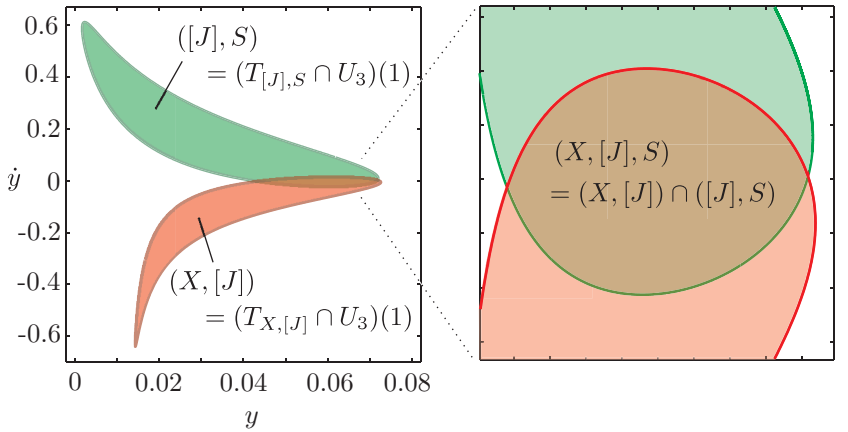


FIGURE 4.4.6. **An itinerary region with label (X,[J],S).** The regions $(X, [J])$ and $([J], S)$ on U_3 are shown to intersect in the left panel. The right panel shows a close-up of the intersection region, $(X, [J], S)$, which contains initial conditons for orbits with itinerary (X, J, S). See the text for details.

1. First, we know from our choice of Poincaré section (U_3) that $x_0 = 1 - \mu$.
2. We then pick values $(y_0, \dot{y}_0) \in (X, [J], S) \subset U_3$
3. Finally, $\dot{x}_0$ comes from the energy equation (2.3.13).

$$\dot{x}_0 = -\sqrt{-\dot{y}_0^2 - 2\bar{U}(x_0, y_0) + 2E} \tag{4.4.4}$$

where the effective potential is given in Eq. (2.3.9). We take the negative sign of the square root by the definition of the U_3 Poincaré section.

We want the solution $s(t)$ which passes through s_0 at time $t = 0$, i.e., $s(0) = s_0$. Evolving our initial point s_0 forward and backward under the equations of motion (2.5.1) within some time interval $[-\tau, \tau]$ for $\tau > 0$ yields the desired solution. We are guided in our choice of τ by the integration times of the trajectories along the tube boundary, which yields an initial guess for τ. Through simple trial and error starting from a reasonable guess, we find the τ which produces the appropriate trajectory, whose projection onto position space looks like that shown in Figure 4.4.1.

4.5 Example Itinerary: (X, J, S, J, X)

In what follows, we apply the numerical construction techniques discussed above to an example with a longer itinerary ($k = 5$). As our example, we

construct a trajectory with itinerary (X, J, S, J, X). This example is chosen because it roughly corresponds to the behavior of comet P/Oterma with respect to the sun-Jupiter system during the years 1910 to 1980 (see Koon, Lo, Marsden, and Ross [2001a]).

We seek itinerary regions with label (X, J, S, J, X) on one of the Poincaré sections $\{U_i\}$. We use the energy, $e = -1.519$, which is in the range $[E_2, E_3]$ for the sun-Jupiter system ($\mu = 9.537 \times 10^{-4}$).

In Figure 4.5.1, we show the first few intersections of the L_1 and L_2 p.o. tubes with U_3. We need to introduce a change in notation. What

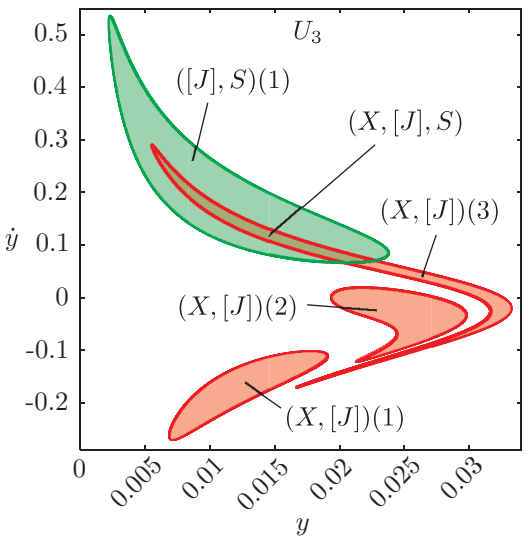


FIGURE 4.5.1. The first intersection of $T_{[J],S}$ and the first three intersections of $T_{X,[J]}$ with U_3 are shown. We use the notation $(X, [J])(n) = (T_{X,[J]} \bigcap U_3)(n)$ for the nth intersection of $T_{X,[J]}$ with U_3. The intersection $(X, [J], S) = (X, [J])(3) \bigcap ([J], S)(1)$ contains all solutions $s(t)$ which come from the X realm, perform two full revolutions around Jupiter in the J realm, and then exit to the S realm.

we called $(X, [J])$ in the previous section is now simply the first intersection of the tube, $T_{X,[J]}$, with the Poincaré section U_3, denoted $(X, [J])(1)$. Similarly, we use the notation $(X, [J])(n) = (T_{X,[J]} \bigcap U_3)(n)$ for the nth intersection of $T_{X,[J]}$ with U_3.

Let

$$f_{33} : U_3 \to U_3,$$
$$(y, \dot{y}) \mapsto (y', \dot{y}'),$$

denote the Poincaré map from U_3 to U_3 (or at least defined on an appropriate restriction of U_3). The map f_{33} is area preserving owing to the Hamiltonian nature of the flow and the choice of Poincaré section. In

particular, we have $(X,[J])(n) = f_{33}((X,[J])(1))$ and $m((X,[J])(n)) = m((X,[J])(n-1))$ where $m(A) \geq 0$ denotes the usual two-dimensional area of a set $A \subset \mathbb{R}^2$.

There is an intersection, $(X,[J],S) = (X,[J])(3) \bigcap ([J],S)(1)$. All initial conditions $s_0 \in (X,[J],S)$ correspond to solutions $s(t)$ which come from the X realm, perform two full revolutions around Jupiter in the J realm, and then exit to the S realm.

To find itinerary regions with the additional symbols, we take the $(X,[J],S)$ region and evolve it forward under the equations of motion (2.5.1) until it intersects the U_1 section in the S realm, shown in Figure 4.5.2. Following our notation, this set is labeled $(X,J,[S])$. Notice that it lies entirely within

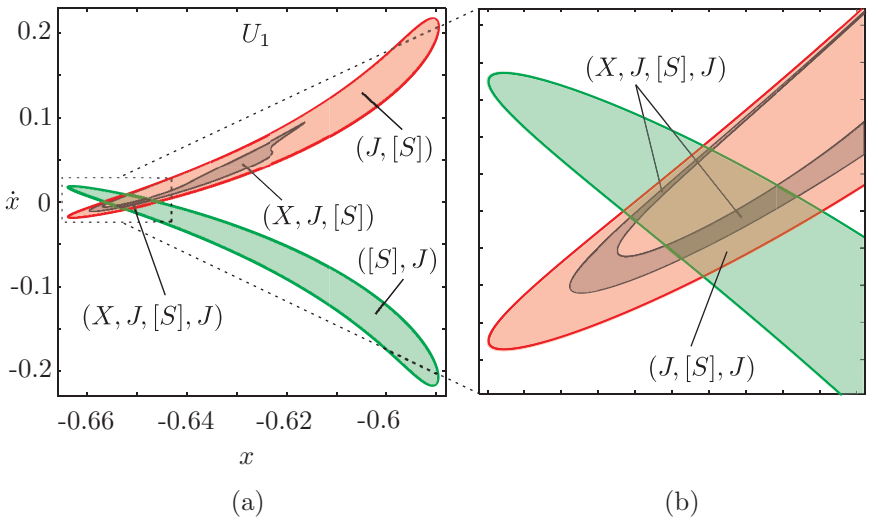


FIGURE 4.5.2. (a) The U_1 Poincaré section in the S realm is shown. $(X,J,[S])$ is obtained by evolving the $(X,[J],S) \subset U_3$ region forward until it intersects U_1. (b) A close-up of the intersection of $(X,J,[S])$ with $([S],J)$, i.e., the $(X,J,[S],J)$ region, which consists of two disconnected large pieces. For this construction, we follow the evolution of the larger piece.

the $(J,[S])$ region, as we would expect. Furthermore, as seen in Figure 4.5.1, part of the boundary of $(X,[J],S)$ is on the boundary of $([J],S)(1)$, i.e., the boundary of the tubes connecting the J and S realms. Due to the infinite winding near the boundary of the tubes upon their approach to the L_1 p.o., this portion of the $(X,J,[S])$ set spirals around the boundary of $(J,[S])$ in U_1, as is suggested in Figure 4.5.2.

The intersection $(X,J,[S]) \bigcap ([S],J)$ consists of a pair of large strips (and infinitely many smaller ones, due to the infinite winding described in Chapter 2), reminiscent of the strips around heteroclinic and homoclinic points which we encountered in Chapter 3. The pair of strips, shown close-up

in Figure 4.5.2(b), carry the label $(X, J.[S], J)$ and bring us one symbol closer to our desired itinerary region. Taking the larger of the two strips, we evolve it forward in time until it re-enters the J realm and intersects U_2, shown in Figure 4.5.3. Notice the symmetry between Figure 4.5.3 and Figure 4.5.1, i.e., $y \mapsto -y, t \mapsto -t$, the symmetry (3.3.1) of the PCR3BP equations of motion.

The $(X, J, S, [J])$ and $(S, [J])$ regions intersect in a thin strip, the desired $(X, J, S, [J], X)$ region. Any trajectory passing through this strip will escape from the J to the X realm in forward time, and will perform a $S \to J \to X$ journey in backward time.

Taking any initial condition in this strip and numerically integrating it forward and backward in time yields a trajectory with the desired itinerary. We give a numerical example in Figure 4.5.4. Orbits in the itinerary region are considered *robust* because nearby orbits have the same itinerary. Regions corresponding to other allowable itineraries of any length can theoretically be generated with this same systematic procedure. Not only do we know such orbits exist from Theorem 3.10.1, but we have a relatively simple method for producing them.

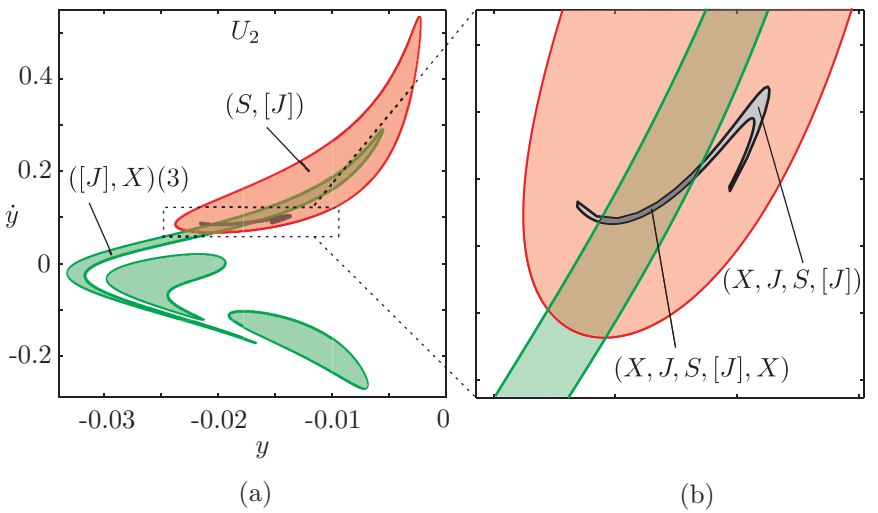


FIGURE 4.5.3. (a) We evolve the larger of the $(X, J, [S], J)$ pieces on U_1 until it intersects U_2. (b) A close-up of the intersection of the $(X, J, S, [J])$ and $([J], X)$ itinerary regions is shown, yielding the desired $(X, J, S, [J], X)$ region.

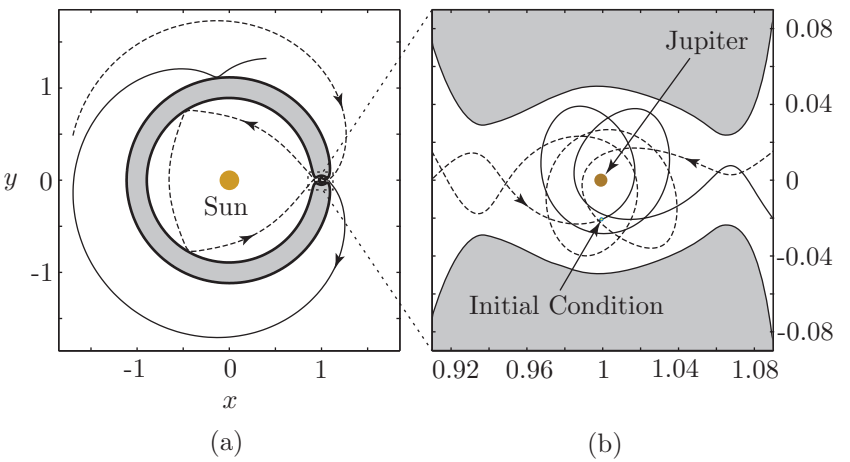


FIGURE 4.5.4. (a) **A trajectory with the itinerary** (X, J, S, J, X) computed using an initial condition inside the $(X, J, S, [J], X)$ region of the U_2 Poincaré section, shown in Figure 4.5.3. (b) A close-up of this trajectory in the J realm is shown and the initial condition is labeled. The backward (respectively, forward) integrated portion of the trajectory is a dashed (respectively, solid) curve.

5
Trajectories in the Four-Body Problem

5.1 Introduction

In this chapter, we describe a procedure to construct trajectories for a spacecraft in the four-body problem using solutions from the three-body problem covered in Chapter 4. We illustrate the procedure in the construction of an important example mission: a low-energy transfer to the Earth's Moon which uses ballistic capture. We will also mention another, closely related example mission concept: transfer between libration point orbits in the Earth-Moon and Sun-Earth systems.

The Patched-Conic Approximation. For many years, trajectory designers for spacecraft on interplanetary missions have obtained good initial trajectory solutions in the N-body problem by dividing the spacecraft's motion into pieces in which the influence of *only one body at a time is considered*. This patched two-body, or patched conic, approximation (see Bate, Mueller, and White [1971]) has worked well for missions such as the Voyager probes which have high relative velocity encounters with the bodies they visit. The criterion for switching from the influence of one body to another involves a dividing surface in the configuration space known as the sphere of influence, taken to be the (nearly) spherical surface about the smaller primary where the acceleration due to both primaries are equal. In nondimensional units, its radius is given approximately (see, e.g., Roy

[1988]) by

$$r_{si} = \mu^{\frac{2}{5}}. \tag{5.1.1}$$

Note that for nearly all μ values of interest in the solar system, this is smaller than the Hill radius encountered in Chapter 2.

The patched conic approximation breaks down when we consider low relative velocity encounters, which are critical for low energy trajectories. In this regime, two bodies (e.g., the Earth and Moon) both influence the motion of the spacecraft with the same order of magnitude, and the restricted three-body problem must be used to model the motion of the spacecraft. Furthermore, the criterion for switching between the influence of one pair of bodies to another pair involves a dividing surface in the full phase space.

Spacecraft Trajectory Design Using the Patched Three-Body Approximation. In this chapter, we describe how to construct trajectories in the four-body problem using invariant manifold tubes from multiple three-body systems using the ***patched three-body approximation***. This is important for taking full advantage of an N-body gravitational field to reduce fuel consumption, and is especially useful in the design of interplanetary trajectories which *visit multiple bodies*. These may include mission trajectories such as a low energy mission to orbit multiple moons of Jupiter or a low energy transfer from the Earth to the Moon. For instance, using the phase space tubes in each three-body system, we are able to construct a transfer trajectory from the Earth which executes an unpropelled (i.e., ballistic) capture at the Moon. An Earth-to-Moon trajectory of this type, which utilizes the perturbation by the Sun, requires less fuel than the usual Hohmann transfer, such as those used by the Apollo missions of the 1960s.

To design, for instance, a spacecraft trajectory from the Earth to the moon which also takes advantage of the sun's gravity, we would want to model the trajectory as two pieces: the first piece being a solution of the sun-Earth-spacecraft system (where the moon's gravitational influence is unimportant) and the second piece being a solution of the Earth-moon-spacecraft system (where the sun's influence is important). The two pieces are connected by two initial conditions, s_{pp}^- and s_{pp}^+, which together form the *patch point* between two three-body solution arcs. Both s_{pp}^- and s_{pp}^+ are at the same location in position space, but we permit them to have differing velocities. The velocity discontinuity, of norm ΔV, corresponds to the impulsive rocket maneuver which will be necessary to effect the transition between the two three-body solutions. Evolving s_{pp}^- backward in time gives the first piece; a solution in one three-body system, e.g., sun-Earth-spacecraft. Evolving s_{pp}^+ forward in time gives the second piece; a solution in the other three-body system, e.g., Earth-moon-spacecraft. We will discuss how to find appropriate patch points pairs, $s_{pp}^{\pm}$.

5.2 Modeling the Four-Body Problem

Consider a particle P in field of three massive bodies, M_0, M_1, and M_2. We suppose that the massive bodies are in one of the two prescribed motions about one another:

1. **Concentric Circular Model (CCM).** M_0 is a central body about which M_1 and M_2 move in circular orbits of radii d_1 and d_2, respectively, where $d_2 > d_1$. In general, we suppose $M_1, M_2 \ll M_0$. This is a model of, e.g., the Jupiter-Ganymede-Callisto system (as M_0, M_1, and M_2, respectively).

2. **Bicircular Model (BCM).** M_1 and M_2 are in circular motion about their barycenter, with mutual separation d_1. Considering all the mass in the M_1-M_2 system to be concentrated at its barycenter, we suppose M_0 and the M_1-M_2 barycenter are in a circular orbit of radius $d_2 > d_1$ about their common center of mass. In general, we suppose $M_2 \ll M_1 \ll M_0$. This is a model of, e.g., the sun-Earth-Moon system (as M_0, M_1, and M_2, respectively).

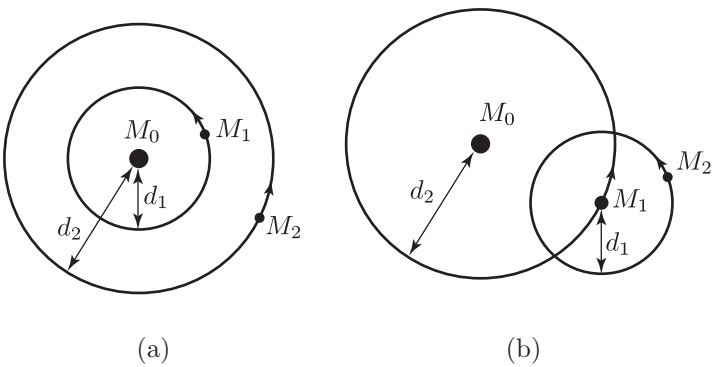


FIGURE 5.2.1. **Models of motion for the four-body problem.** (a) *Concentric Circular Model.* M_0 is a central body about which M_1 and M_2 move in circular orbits of radii d_1 and d_2, respectively, where $d_2 > d_1$. (b) *Bicircular Model.* M_1 and M_2 are in circular motion about their barycenter, with mutual separation d_1. Considering all the mass in the M_1-M_2 system to be concentrated at its barycenter, we suppose M_0 and the M_1-M_2 barycenter are in a circular orbit of radius $d_2 > d_1$ about their common center of mass.

5.3 Bicircular Model

In this chapter, we will only consider the BCM, which can be used to compute a low-energy Earth-to-Moon trajectory or Earth to lunar libration point trajectory. The CCM has been used to produce a "Petit Grand Tour" of two of Jupiter's moons, as reported in Koon, Lo, Marsden, and Ross [1999] and Gómez, Koon, Lo, Marsden, Masdemont, and Ross [2001]. More recently, an extension of the CCM has been used to design a "Multi-Moon Orbiter" of three of Jupiter's moons (see Ross, Koon, Lo, and Marsden [2003, 2004]), whose discussion we defer to Chapter 10.

Equations of Motion in Earth-Moon Rotating Frame As mentioned earlier, we use the equations of motion derived under the BCM assumptions as the underlying dynamical model. The bicircular problem is a simplified version of the restricted four-body problem. The objective is to describe the motion of a spacecraft of negligible mass under the gravitational attraction of the Earth, Moon, and Sun. "Negligible mass" means that the spacecraft does not influence the motion of the Earth, Moon, and Sun. This description follows that of Simó, Gómez, Jorba, and Masdemont [1995].

In this model we suppose that the Earth and Moon are revolving in circular orbits around their center of mass (barycenter) and the Earth-Moon barycenter is moving in a circular orbit around the center of mass of the Sun-Earth-Moon system. The orbits of all four bodies are in the same plane. We remark that, with these assumptions, the motion of these three bodies is not coherent. That is, the assumed motions do not satisfy Newton's equations. However, numerical simulation shows that, in some regions of phase space, this model gives the same qualitative behavior as the real system. Thus, the model is extremely useful for the study of some kinds of orbits, in particular the Hiten trajectory of Belbruno and Miller [1993] and more recently the "Shoot the Moon" trajectory of Koon, Lo, Marsden, and Ross [2001b].

To simplify the equations, the units of length, time, and mass are chosen such that the angular velocity of rotation of the Earth and Moon (around their barycenter), the sum of the masses of the Earth and Moon, and the gravitational constant are all equal to one. With these normalized units, the Earth-Moon distance is also one. Let μ be the mass of the Moon in these units. Then $1-\mu$ the mass of the Earth. Let m_S the mass of the Sun. Let the semimajor axis of the Sun be a_S.

We use a synodic (rotating) coordinates with respect to the Earth-Moon system. The origin is taken at the center of mass of the Earth-Moon system. The x-axis is given by the line that goes from the Earth to the Moon, and the y-axis is taken such that the system is orthogonal and positive oriented. Note that, in this synodic (non-inertial) frame, the Earth and Moon have fixed positions and the Sun is rotating clockwise around the barycenter of the Earth-Moon system. The positions of the Earth and Moon are fixed at $(-\mu, 0)$ and $(1-\mu, 0)$, respectively. The angular velocity of the Sun in

these synodic coordinates is denoted by ω_S and the phase of the Sun at $t = 0$ is θ_{S0}. See Figure 5.3.1.

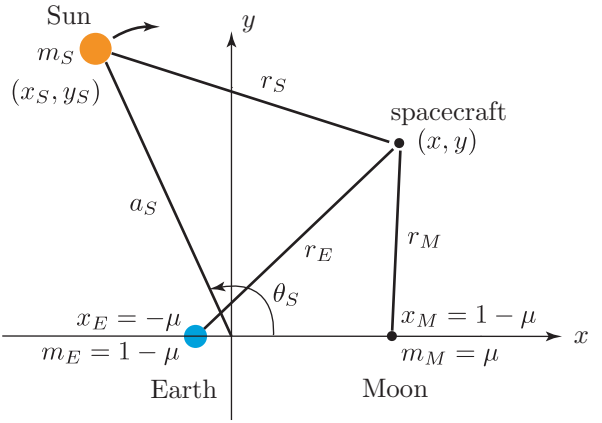


FIGURE 5.3.1. Rotating coordinate frame in the BCM approximation with Earth and Moon fixed on the x-axis. As seen in this frame, the Sun rotates clockwise around the Earth-Moon barycenter (the origin) with angular frequency ω_S.

Using nondimensional units, the equations of motion in the BCM are

$$\begin{aligned}
\dot{x} &= u,\\
\dot{y} &= v,\\
\dot{u} &= x + 2v - c_E(x + \mu_M) - c_M(x - \mu_E) - c_S(x - x_S) - \alpha_S x_S,\\
\dot{v} &= y - 2u - c_E y - c_M y - c_S(y - y_S) - \alpha_S y_S,
\end{aligned} \tag{5.3.1}$$

where

$$c_i = \frac{\mu_i}{r_i^3}, \quad \text{for} \quad i = E, M, S, \tag{5.3.2}$$

$$\alpha_S = \frac{m_S}{a_S^3}, \tag{5.3.3}$$

and

$$\begin{aligned}
r_E &= \sqrt{(x + \mu_M)^2 + y^2},\\
r_M &= \sqrt{(x - \mu_E)^2 + y^2},\\
r_S &= \sqrt{(x - x_S)^2 + (y - y_S)^2},
\end{aligned}$$

with

$$\begin{aligned}\mu_E &= 1-\mu,\\ \mu_M &= \mu,\\ x_S &= a_S\cos(\theta_S),\\ y_S &= a_S\sin(\theta_S),\\ \theta_S &= -\omega_S t+\theta_{S0}.\end{aligned}$$

The values of the parameters are as follows:

$$\mu=\frac{m_M}{m_M+m_E}=0.01215, \tag{5.3.4}$$

from Table 2.2.1 in Chapter 2 where m_M and m_E denote the mass of the Moon and Earth, respectively. The Sun's mass (1 unit = Earth + Moon mass) is

$$m_S=328900.54, \tag{5.3.5}$$

the Sun's distance (1 unit = Earth-Moon distance) is

$$a_S=388.81114, \tag{5.3.6}$$

and the Sun's angular velocity in synodic coordinates is

$$\omega_S=0.925195985520347. \tag{5.3.7}$$

In the above equations, time is scaled by the period of the Earth and Moon around their center of mass ($T/2\pi$, where $T=2.361\times10^6$ s), positions are scaled by the average Earth-Moon distance ($L=3.850\times10^5$ km), and velocities are scaled by the Moon's average orbital speed around the Earth ($2\pi L/T=1.025$ km/s). One can find these values in Table 2.2.1.

Equations of Motion in Sun-Earth Rotating Frame. In this model we suppose that the Sun and Earth are revolving in circular orbits around their barycenter and the Moon is moving in a circular orbit around the center of the Earth. The orbits of all four bodies are in the same plane. We remark that, with these assumptions, the motion of these three bodies is not coherent. That is, the assumed motions do not satisfy Newton's equations. But as this model is extremely useful for the study of some kinds of orbits, we will use it.

Let μ be the mass of the Earth, $1-\mu$ the mass of the Sun and m_M the mass of the Moon. Let the distance between the Sun and the Earth be taken as unity. Let the orbit of the Sun and Earth around the Sun-Earth barycenter also be taken as unity. The distance from the Earth to the Moon is a_M. We use rotating coordinates with respect to the Sun-Earth system, so that the positions of the Sun and Earth are fixed at $(-\mu,0)$ and $(1-\mu,0)$, respectively. The angular velocity of the Moon in these synodic coordinates is denoted by ω_M and the phase of the Moon at $t=0$ is θ_{M0}.

In the rotating frame just defined and using nondimensional units, the equations of motion in the Sun-Earth rotating frame are

$$\begin{aligned}
\dot{x} &= u,\\
\dot{y} &= v,\\
\dot{u} &= x + 2v - c_S(x+\mu_E) - c_E(x-\mu_S) - c_M(x-x_M),\\
\dot{v} &= y - 2u - c_S y - c_E y - c_M(y-y_M),
\end{aligned} \tag{5.3.8}$$

where

$$c_i = \frac{\mu_i}{r_i^3}, \quad \text{for} \quad i = S, E, M \tag{5.3.9}$$

$$\alpha_S = \frac{m_S}{a_S^3}, \tag{5.3.10}$$

and

$$\begin{aligned}
r_S &= \sqrt{(x+\mu_E)^2 + y^2},\\
r_E &= \sqrt{(x-\mu_S)^2 + y^2},\\
r_M &= \sqrt{(x-x_M)^2 + (y-y_M)^2},
\end{aligned}$$

with

$$\begin{aligned}
\mu_S &= 1 - \mu,\\
\mu_E &= \mu,\\
x_M &= a_M \cos(\theta_M),\\
y_M &= a_M \sin(\theta_M),\\
\theta_M &= \omega_M t + \theta_{M0}.
\end{aligned}$$

The values of the parameters are as follows:

$$\mu = \frac{m_E}{m_E + m_S} = 3.036 \times 10^{-6}, \tag{5.3.11}$$

where m_E and m_S denote the mass of the Earth and Sun, respectively. The Moon's mass (1 unit = Sun + Earth mass) is

$$m_M = 3.733998734625702 \times 10^{-8}. \tag{5.3.12}$$

the Earth-Moon distance (1 unit = Sun-Earth) is

$$a_M = 2.573565073532068 \times 10^{-3}, \tag{5.3.13}$$

and the Moon's angular velocity in synodic coordinates is

$$\omega_M = 12.36886949284508. \tag{5.3.14}$$

In the above equations, time is scaled by the period of the Sun and Earth around their center of mass ($T/2\pi$, where $T = 3.156 \times 10^7$ s), positions are scaled by the average Sun-Earth distance ($L = 1.496 \times 10^8$ km), and velocities are scaled by the Earth's average orbital speed around the Sun ($2\pi L/T = 29.7840$ km/s), according to Table 2.2.1.

Transforming Coordinates between Rotating Frames. A necessary algorithm is the transformation between the two rotating coordinate frames: the Earth-Moon rotating frame and the Sun-Earth rotating frame.

Let the phase space trajectory in rotating coordinate system A be denoted by $\mathbf{x}_A^{\text{ro}}(t_A)$ where $\mathbf{x}_A^{\text{ro}} = [x, y, u, v]^T$ is in the nondimensional position and velocity units associated with system A and t_A is in the corresponding time units of system A.

We first transform to inertial coordinates centered on the primary m_i, $i = 1$ or 2, via

$$\mathbf{x}_A^{\text{in}} = \mathbf{R}(\mathbf{x}_A^{\text{ro}} - \mathbf{d}_A), \tag{5.3.15}$$

where

$$\mathbf{R} = \begin{pmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{pmatrix}, \tag{5.3.16}$$

$$R_{11} = R_{22} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}, R_{21} = \begin{pmatrix} -s & -c \\ c & -s \end{pmatrix}, \tag{5.3.17}$$

$$c = \cos(\theta(t_A)), s = \sin(\theta(t_A)),$$
$$\theta(t_A) = t_A + \theta_{A0},$$

and $\mathbf{d}_A = [x_A^0, 0, 0, 0]^T$ and x_A^0 is $-\mu_A$ or $1 - \mu_A$ depending on whether the A system inertial frame is m_1- or m_2-centered, respectively.

We then change from the units of system A to the units of another system, B. Let $L_{AB} = \frac{L_A}{L_B}$ be the ratio of the length scales and $T_{AB} = \frac{T_A}{T_B}$ be the ratio of the time scales. The inertial frame position, velocity, and time coordinates in the B system are then given by the re-scaling,

$$x_B^{\text{in,pos}} = L_{AB} x_A^{\text{in,pos}}, \tag{5.3.18}$$

$$x_B^{\text{in,vel}} = \frac{L_{AB}}{T_{AB}} x_A^{\text{in,vel}}, \tag{5.3.19}$$

$$t_B = T_{AB} t_A, \tag{5.3.20}$$

respectively.

The primary m_i of system A is the primary m_j of system B. Thus, the trajectory $\mathbf{x}_B^{\text{in}}(t_B)$ is in m_j-centered inertial coordinates in the units of system B. To transform back to rotating coordinates, we use

$$\mathbf{x}_B^{\text{ro}} = \mathbf{R}^{-1} \mathbf{x}_B^{\text{in}} + \mathbf{d}_B, \tag{5.3.21}$$

where $\mathbf{d}_B = [x_B^0, 0, 0, 0]^T$ and x_B^0 is $-\mu_B$ or $1 - \mu_B$ depending on whether the B system inertial frame is m_1- or m_2-centered, respectively.

5.4 Example 1: Low-Energy Transfer to the Moon

Hiten Mission. The traditional approach to construct a spacecraft transfer trajectory to the moon from the Earth is by Hohmann transfer. This type of transfer uses only two-body dynamics. It is constructed by determining a two-body Keplerian ellipse from an Earth parking orbit to the orbit of the moon, illustrated schematically in Figure 5.4.1(a). The two bodies involved are the Earth and a spacecraft. Such a transfer requires a large ΔV for the spacecraft to get captured by the moon.

In 1991, the failed Japanese mission, Muses-A, whose propellant budget did not permit it to transfer to the moon via the usual method was given a new life with an innovative trajectory design, based on the work of Belbruno and Miller [1993]. Its re-incarnation, renamed Hiten, used a low-energy transfer with a ballistic capture at the moon. An Earth-to-Moon trajectory of this type, shown in Figures 5.4.1(b) and (c), which utilizes the perturbation by the Sun, requires less fuel than the usual Hohmann transfer.

Using the Patched Three-Body Approximation to Systematically Design Earth-to-Moon Trajectories with Ballistic Capture. In this section, we present an approach to the problem of the orbital dynamics of this interesting trajectory by implementing the view that the Sun-Earth-Moon-spacecraft four-body system can be approximated as two three-body systems (a view also taken by Belbruno [1994]). Figure 5.4.2(a) shows a schematic of this trajectory in the Sun-Earth rotating frame, showing the two legs of the trajectory: (1) the Sun-Earth libration point portion and (2) the lunar capture portion.

Within each three-body system, using our understanding of the invariant manifold structures associated with the libration points L_1 and L_2, we transfer from a 200 km altitude Earth orbit into the region where the invariant manifold structure of the Sun-Earth libration points interact with the invariant manifold structure of the Earth-Moon libration points. See Figure 5.4.2(b). We utilize the sensitivity of the "twisting" of trajectories near the invariant manifold tubes in the libration point region to find a fuel efficient transfer from the Sun-Earth system to the Earth-Moon system. The invariant manifold tubes of the Earth-Moon system provide the dynamical channels in phase space that enable ballistic captures of the spacecraft by the Moon.

The final Earth-to-Moon trajectory is integrated in the bicircular four-body model described in §5.3, where both the Moon and the Earth are assumed to move in circular orbits about the Earth and the Sun, respectively, in the ecliptic, and the spacecraft is an infinitesimal mass point.

The success of this approach depends greatly on the configuration of the specific four bodies of interest. In order for low-energy transfers to take place, the invariant manifold structures of the two three-body systems must intersect within a reasonable time. Otherwise, the transfer may require an impractically long time of flight. For the Sun-Earth-Moon-spacecraft case,

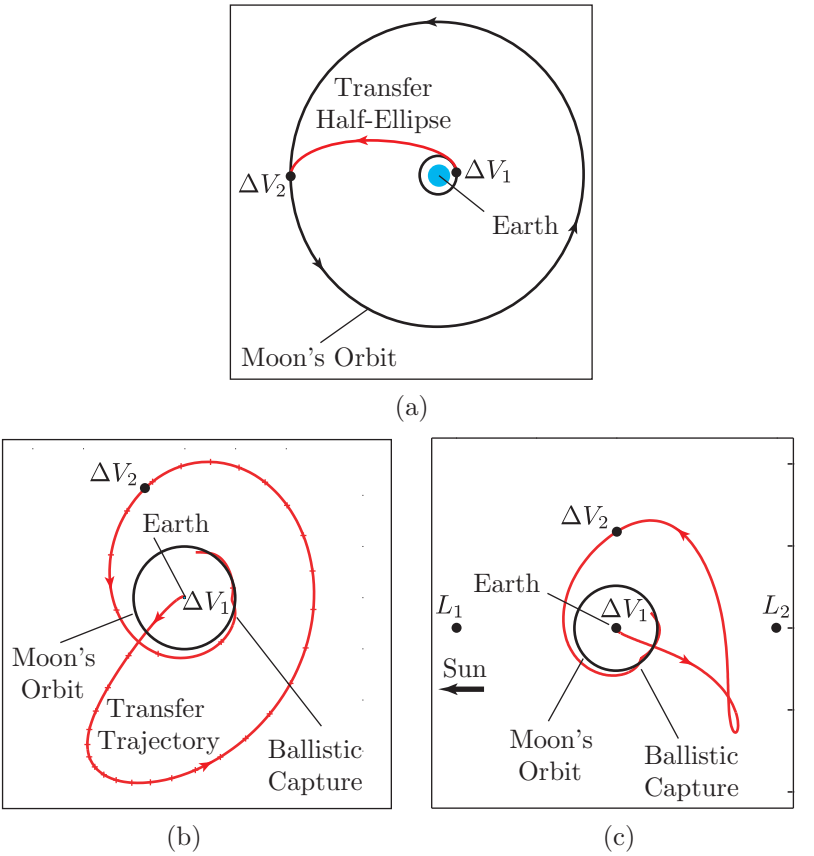


FIGURE 5.4.1. (a) Hohmann transfer. (b) Low-energy transfer trajectory in the geocentric inertial frame. (c) Same trajectory in the Sun-Earth rotating frame.

this is not a problem. The overlap of these invariant manifold structures provide the low-energy transfers between the Earth and the Moon.

Construction of Earth-to-Moon Transfer. The construction is done mainly in the Sun-Earth rotating frame using the Poincaré section Γ (along a line of constant x-position passing through the Earth). This Poincaré section helps to glue the Sun-Earth libration point portion of the trajectory with the lunar ballistic capture portion.

The basic strategy is to find an initial condition (position and velocity) for a spacecraft on the Poincaré section such that when integrating forward, the spacecraft will be guided by the L_2 Earth-Moon manifold and get ballistically captured by the Moon; when integrating backward, the spacecraft will hug the Sun-Earth manifolds and return to Earth.

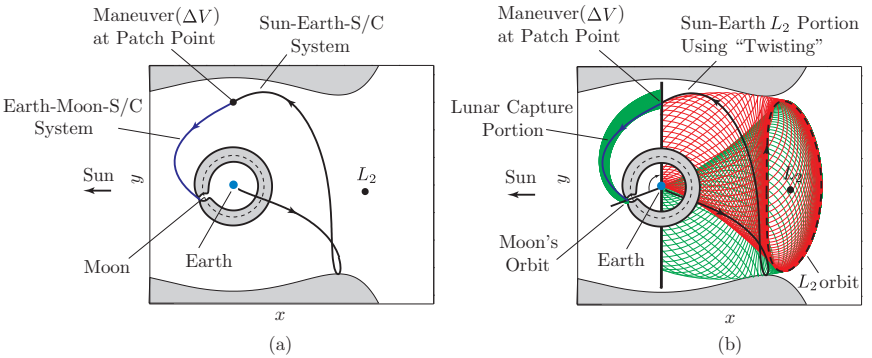


FIGURE 5.4.2. (a) Two legs of a Hiten-like trajectory in the Sun-Earth rotating frame. (b) The interaction of invariant manifold tubes of the Sun-Earth and the Earth-Moon systems permits a fuel efficient Earth-to-Moon transfer with the perturbation of the Sun.

We utilize two important properties of the libration point dynamics of the three-body problem. The stable manifold tube is key in targeting a capture orbit for the Earth-Moon portion of the design. The twisting of orbits in the equilibrium region is key in finding a fuel efficient transfer for the Sun-Earth libration point portion of the trajectory.

Lunar Ballistic Capture Portion. Recall that by targeting the region enclosed by the stable manifold tube of the L_2 Lyapunov orbit in the Earth-Moon system, we can construct an orbit which will get ballistically captured by the Moon. When we transform this Poincaré cut of the stable manifold of an Earth-Moon L_2 Lyapunov orbit into the Poincaré section of the Sun-Earth system, we obtain a closed curve. A point interior to this curve will approach the Moon when integrated forward. See Figure 5.4.3. Assuming the Sun is a negligible perturbation to the Earth-Moon-spacecraft three-body dynamics during this leg of the trajectory, any spacecraft with initial conditions within this closed curve will be ballistically captured by the Moon. "Ballistic capture by the Moon" means an orbit which under natural dynamics gets within the Hill radius of the Moon (approx. 60,000 km) and performs at least one revolution around the Moon. In such a state, a slight ΔV will result in a stable capture (closing off the necks at L_1 and L_2).

Twisting of Orbits and Sun-Earth libration Point Portion. Since the twisting of orbits in the equilibrium region is key in finding the Sun-Earth libration point portion of the design, we would like to review this property briefly. From Chapter 2, we learn that orbits twist in the equilibrium region following roughly the Lyapunov orbit. The amount of twist of an orbit depends sensitively on its distance from the manifold tube. The closer to the manifold tube an orbit begins on its approach to the equilibrium region, the more it will be twisted when it exits the equilibrium

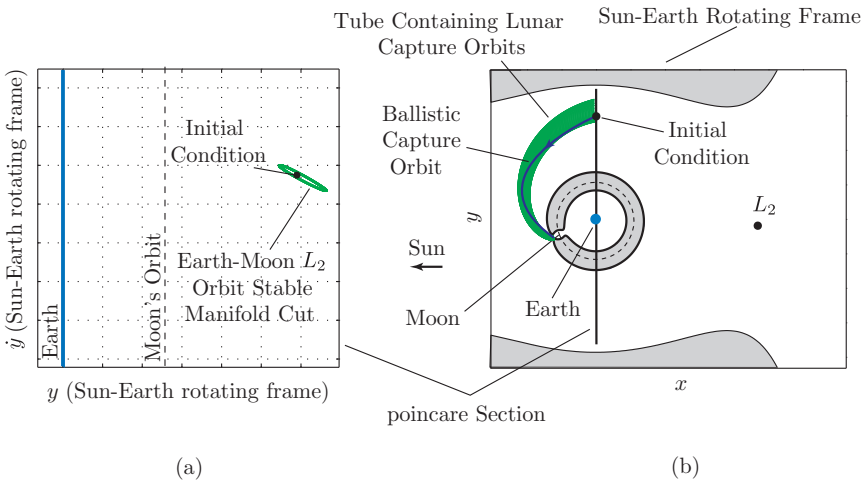


FIGURE 5.4.3. (a) The stable manifold cut of an Earth-Moon L_2 orbit in the Poincaré section of the Sun-Earth system. (b) A point interior to this cut, with the correct phasing of the Moon, will reach the Moon's ballistic capture region when integrated forward.

region. Hence, with small change in the initial condition (such as a small change in velocity at a fixed point), we can change the destination of an orbit dramatically. In fact, we can use this sensitivity to target the spacecraft back to a 200 km Earth parking orbit.

Look at the Poincaré section Γ in Figure 5.4.4(a). Notice that a strip q_2q_1 of orbits just outside of the unstable manifold cut, when integrated backward, gets stretched into a long strip $P^{-1}(q_2)P^{-1}(q_1)$ of orbits that wraps around the whole stable manifold cut. Points on q_2q_1 represent orbits which have the same position but slightly different velocity. But their preimage $P^{-1}(q_2)P^{-1}(q_1)$ can reach any position on the lower line of Figure 5.4.4(b) where the stable manifold tube intersects.

Pick an energy in the temporary capture range of the Sun-Earth system which has L_2 orbit manifolds that come near a 200 km altitude Earth parking orbit. Compute the Poincaré section Γ (see Figure 5.4.4(a)). The curve on the right is the Poincaré cut of the unstable manifold of the Lyapunov orbit around the Sun-Earth L_2. Picking an appropriate initial condition just outside this curve, we can backward integrate to produce a trajectory coming back to the Earth parking orbit.

Connecting the Two Portions. We can vary the phase of the Moon until the Earth-Moon L_2 manifold cut intersects the Sun-Earth L_2 manifold cut, as illustrated in Figures 5.4.5(a) and (b). In the region which is in the interior of the Earth-Moon L_2 manifold curve but in the exterior of the Sun-Earth L_2 manifold curve, an orbit will get ballistically captured by the Moon when integrated forward; when integrated backward, the orbit

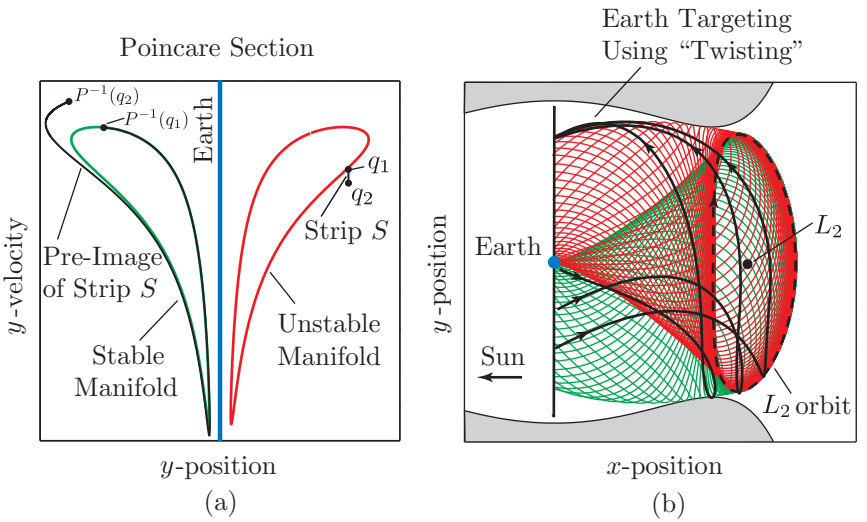


FIGURE 5.4.4. (a) Line strip q_2q_1 outside of unstable manifold cut gets stretched into a long strip $P^{-1}(q_2)P^{-1}(q_1)$ that wraps around stable manifold cut. (b) With infinitesimal changes in velocity, any point near lower tube cross section can be targeted (integrating backward).

will hug the unstable manifold back to the Sun-Earth L_2 equilibrium region with a twist, and then hug the stable manifold back towards the position of the Earth parking orbit. See Figures 5.4.5(c) and (d).

With only a slight modification (a small mid-course ΔV of 34 m/s at the patch point), this procedure produces a genuine solution integrated in the bicircular four-body problem. Since the capture at the Moon is natural (zero ΔV), the amount of on-board fuel necessary is lowered by about 20% compared to a traditional Hohmann transfer (the Hohmann transfer value is taken from Belbruno and Miller [1993]).

Aside: Why Does It Work?

What follows are a couple of heuristic arguments for using the patched three-body approximation. When outside the Moon's small Hill sphere (60,000 km), which is most of the pre-capture flight, we can consider the Moon's perturbation on the Sun-Earth-spacecraft three-body system to be negligible. Thus, we can utilize Sun-Earth libration point invariant manifold structures. The mid-course ΔV is performed at a point where the spacecraft is re-entering the Earth's sphere of influence (925,000 km), where we can consider the Sun's gravitational perturbation on the Earth-Moon-spacecraft three-body system to be negligible. Thus, Earth-Moon libration point structures can be utilized for the lunar portion of the trajectory.

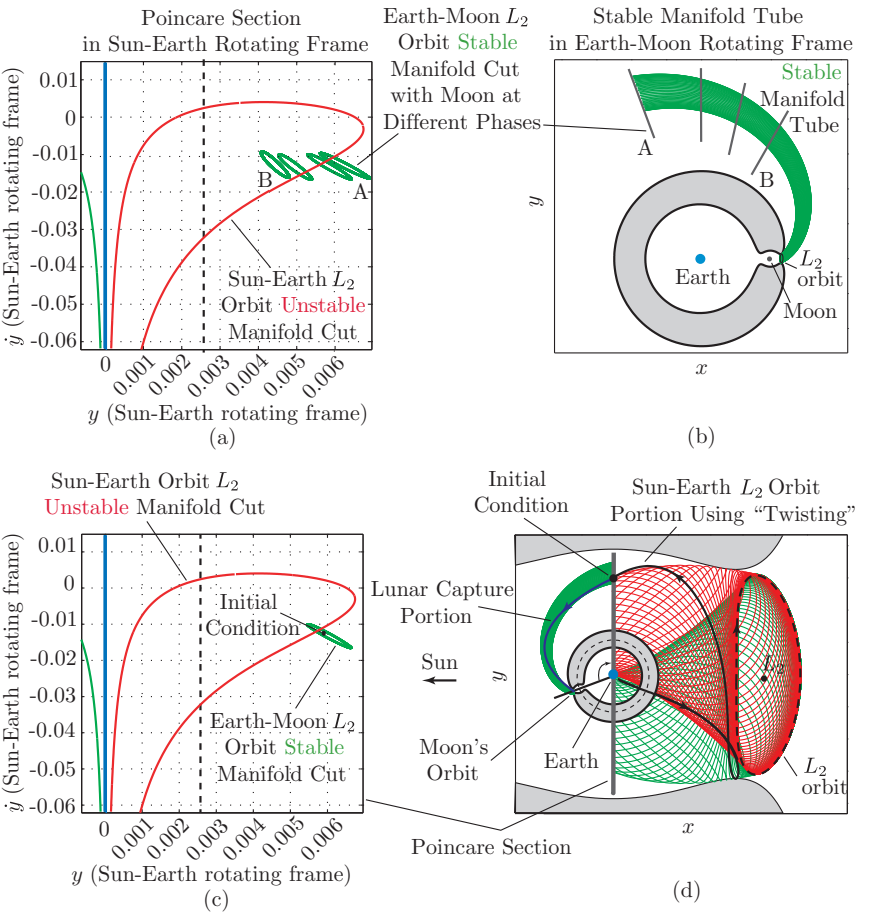


FIGURE 5.4.5. (a) and (b) Vary the phase of the Moon until Earth-Moon L_2 manifold cut intersects Sun-Earth L_2 manifold cut. (c) Pick a point in the interior of the Earth--Moon L_2 manifold curve but in the exterior of the Sun-Earth L_2 manifold curve. (d) An orbit will get ballistically captured by the Moon when integrated foreward; when integrated backward, orbit will hug the invariant manifolds back to the Earth.

Moreover, the fact that the patch point ΔV is so small and may even be eliminated can be understood by considering the following. From a 200 km circular orbit around the Earth, it requires approximately 3150 m/s (provided by the launch vehicle) to reach the Earth-Moon L_1 and L_2. For another 50 m/s, one can reach the Sun-Earth L_1 and L_2. In other words, a spacecraft needs roughly the same amount of energy to reach the Sun-Earth and the Earth-Moon L_1 and L_2. This fortuitous coincidence is what enables these low energy lunar transfer and capture orbits.

5.5 Example 2: The Lunar L_1 Gateway Station

Introduction. NASA desires to develop a robust and flexible capability to visit several potential destinations. Therefore, a Lunar Gateway Station at the lunar L_1 has been proposed as a transportation hub to get humanity beyond low-Earth orbit (see Lo and Ross [2001] and Condon and Pearson [2001]). The fortuitous arrangement of the stable and unstable manifold tubes in near-Earth space implies that lunar L_1 and L_2 periodic orbits are connected to periodic orbits around Earth's L_1 and L_2 via low energy pathways. Many of NASA's future space telescopes located around the Earth's L_1 or L_2 may be built in a lunar L_1 orbit and conveyed to the final destination with minimal fuel requirements. As this methodology works in both the planar and spatial restricted four-body problems (see Chapters 6 and 9), the periodic orbits could be either planar Lyapunov orbits or three-dimensional halo orbits.

Considering the example from the previous section of an Earth to Moon trajectory, suppose that instead of traveling from Earth to the realm around the Moon by entering the *interior* of the Moon's L_2 stable tube, a spacecraft got *on* the Moon's L_2 stable tube. The spacecraft would tend asymptotically to a periodic orbit about the Moon's L_2. We can further suppose that a piece of the Moon's L_1 stable tube, at the same energy, which escapes away from the Moon (in backward time) through the interior of the Moon's L_2 stable tube may intersect the region around Earth's L_2 unstable tube shown in Figure 5.4.5(c). This would be a free transfer from the Earth to a periodic orbit around the Moon's L_1. If the arrow of time were reversed, then the stable and unstable tubes would interchange and one would have a transfer from a periodic orbit about the Moon's L_1 to the Earth. By choosing nearby initial conditions on the Moon's L_1 unstable tube which intersect the Earth's L_2 stable tube, we would have a free transfer from an orbit about the Moon's L_1 to the Earth's L_2, as illustrated in Figure 5.5.1 (the reverse transfer would also exist).

Such a transfer could prove very useful to the development of space. For example, this implies that lunar L_1 halo orbits are connected to halo orbits around Earth's L_1 or L_2 via low energy pathways. As mentioned, future space observatories located around the Earth's L_1 or L_2 may be built in a lunar L_1 orbit and conveyed to the final destination using tubes with minimal propulsion requirements. Similarly, when the spacecraft or instruments require servicing, they may be returned from Earth libration orbits to the lunar L_1 orbit where human servicing may be performed, which was shown to be of vital importance for keeping the Hubble Space Telescope operable. Since the lunar L_1 orbit may be reached by astronauts from Earth in a few days, the infrastructure and complexity of long-term

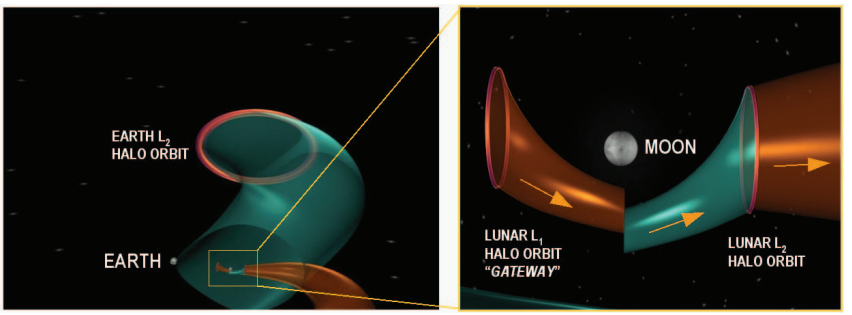


FIGURE 5.5.1. (a) Artist's conception of portions of the stable and unstable tubes of the Sun-Earth-Moon system generated by the halo orbits. The stable tube asymptotically approaching a Earth L_2 halo orbit is shown along with tubes in the Earth-Moon system. (b) An exploded view of the lunar portion of (a). Arrows indicate the direction of movement.

space travel is greatly mitigated. The same orbit could reach any point on the surface of the Moon within hours, thus this portal is also a perfect location for the return of human presence on the Moon. The lunar L_1 orbit is also an excellent point of departure for interplanetary flight where several lunar and Earth encounters may be added to further reduce the launch cost and open up the launch period. The lunar L_1 is a versatile hub for a space transportation system of the future.

Human Servicing of Libration Point Missions from the Lunar L_1. In recent years, halo orbits around the Earth's libration points (EL_1, EL_2, see Figure 5.5.2) have become a popular location for space missions (see Chapter 1). NASA has a lot of experience with halo orbit missions. In 2001

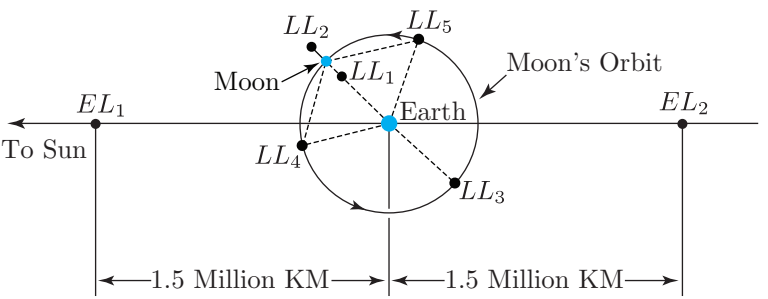


FIGURE 5.5.2. Schematic diagram of the libration points of the Earth-Moon, and Sun-Earth systems.

alone NASA is sending two missions to orbit the Earth's libration points: MAP is well on its way to EL_2 as is Genesis to EL_1. The Next Generation

Space Telescope and the Terrestrial Planet Finder mission (TPF) are both considering using EL_2 orbits. The constant cold environment of EL_2 is well suited to observatories with detectors requiring low temperatures for operation. Communications geometry from EL_2 to the Earth is nearly constant with the range at roughly 1.5 million km from the Earth. Furthermore, it requires a ΔV of only 3200 m/s to insert into typical halo orbits from a 200 km parking orbit around the Earth. In general, operations costs are low: only four to six maneuvers per year are required for station keeping with a total ΔV budget less than 5 m/s per year.

In the last few years, NASA planners have seriously considered providing human servicing to libration missions (see Condon and Pearson [2001]). The problem is that, the 3200 m/s transfer to orbits around the libration points require approximately 3 months of travel time. With transfer orbits to EL_2 well outside of the Earth's magnetic field, such a voyage would in principle be not very different from one going to Mars. To reduce the transfer time in any significant manner (down to one day) requires an increase of the transfer ΔV by roughly an order of magnitude. The infrastructure cost and risk for both options are extremely high. Lo and Ross [2001] suggested an alternate approach by using the Moon's L_1 (lunar L_1: LL_1) as a base of operations for servicing missions at the Earth's libration points.

By placing a Lunar Gateway Habitat in orbit around LL_1, the spacecraft at EL_2 can be brought back and forth to LL_1 with relatively little cost. An example trajectory is presented in this section which requires only a single 14 m/s deterministic maneuver (statistical maneuvers not included) to convey a spacecraft from LL_1 to EL_2 orbit (see Figure 5.5.3). Transfers for EL_2 to LL_1 would have similar costs. With optimization, even this small deterministic maneuver may be removed in some instances. The transfer from the LL_1 to EL_2 region requires about 40 days. This efficient transfer is achieved using a methodology similar to the one in the previous section, with the main difference that the transfer is an intersection of the surface of tubes in the Earth-Moon and Sun-Earth systems, rather than the transit and non-transit orbits on either side of the tube surfaces.

The Lunar L_1 as a Servicing and Transportation Hub. Lunar L_1 is an ideal and logical next step for extended human presence in space beyond LEO (Low Earth Orbit). To first order, from energy considerations, it requires only a ΔV of 3150 m/s to reach LL_1 from a 200 km parking orbit around Earth. Although, this will vary depending on the transfer time. In the worst case, it is bounded above by transfers to the Moon. We are currently studying this issue. Station keeping is required once or twice a week with a total ΔV budget of around 10 m/s per year (Gómez, Howell, Masdemont, and Simó [1998a]). However, advances in navigation technology in the next decade may provide a completely autonomous system for station keeping with even lower cost. Communications is relatively simple, since LL_1 is close by and always in view of the Earth. And, of course, NASA

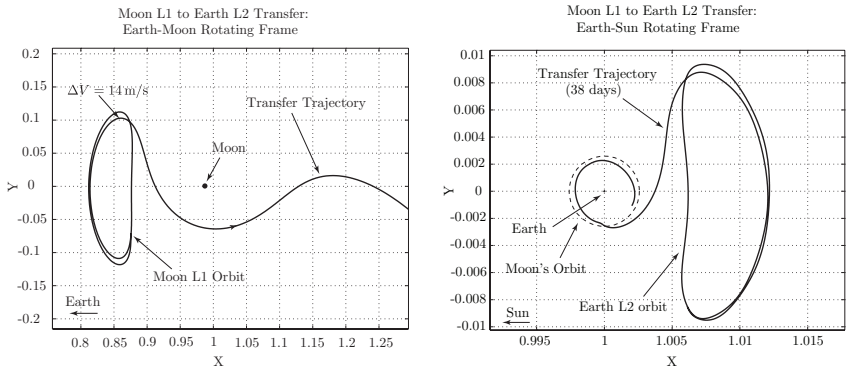


FIGURE 5.5.3. A transfer between planar Lyapunov orbits around the lunar L_1 and Earth L_2. (a) The Lyapunov orbit around the lunar LL_1 and the 14 m/s maneuver to get onto the transfer orbit. The axes are in nondimensional units of the Earth-Moon system. (b) The transfer orbit going from the Moon to the Earth's EL_2. The axes are in nondimensional units of the Sun-Earth system.

has a tremendous amount of experience with human missions to the Moon. This fact alone greatly reduces the risk of this approach.

These facts combine to suggest that a halo orbit around LL_1 provides an ideal location for a "service station" or a "hub" for missions in Earth libration orbits. Moreover, as shown in Doedel, Paffenroth, Keller, Dichmann, Galan, and Vanderbauwhede [2003], there are large families of orbits with similar characteristics to halo orbits in the Earth's Neighborhood (the region between EL_1 and EL_2) which will be useful for future missions. Spacecraft in these orbits may also be serviced by the LL_1 Gateway. Beyond the Earth's Neighborhood, LL_1 can also serve as an excellent point of departure and arrival for interplanetary flights to Mars, the asteroids, and the outer solar system, as depicted in Figure 1.2.5. By taking advantage of the tube dynamics in the Sun-Earth-Moon system, launches from the LL_1 Gateway can effectively increase the narrow launch periods of interplanetary missions from a few days to weeks and months. This is achieved by launching earlier and spending the extra time in the Earth's Neighborhood until a final Earth flyby with injection onto the desired interplanetary transfer orbit. During the time in the Earth's Neighborhood, additional lunar and Earth flybys can further increase the energy of the spacecraft.

6

Extension to the Spatial Problem: Halo Orbits and Their Computation

6.1 Introduction

Extension to Three Degrees-of-Freedom. Earlier chapters focus on the two degree-of-freedom problem, but in this chapter we begin to discuss the three degree-of-freedom circular restricted three-body problem (CR3BP), also referred to as the *spatial problem*. When a comet or spacecraft is relatively far from a planet or moon, the planar model is a good approximation. As one approaches a planet or moon, the dynamics of the third dimension become important. With the addition of a third degree of freedom, new phenomena emerge. In particular, the orbital taxonomy near the libration points gets more complicated.

For space missions, the third dimension is important for geometric reasons such as constant access to the sun for solar power or to the Earth for communications. Also, one is often required to approach a planet or moon out of plane. Controlling the latitude and longitude of a spacecraft's escape from and entry into a planetary or moon orbit require three-dimensional capabilities. For example, a future mission to send a probe to orbit Europa may desire a capture into a high inclination polar orbit around Europa (Sweetser, Maddock, Johannesen, Bell, Penzo, Wolf, Williams, Matousek, and Weinstein [1997]; Ludwinski, Guman, Johannesen, Mitchell, and Staehle [1998]; Scheeres, Guman, and Villac [2001]; Villac and Scheeres [2001]). Three-dimensional capability is also required when decomposing an N-body system into three-body subsystems which are not co-planar, such as the Earth-Sun-spacecraft and Earth-Moon-spacecraft systems. (The tilt

in the orbital planes of the Earth around the Sun and the Moon around the Earth is about 5 degrees.) These demands necessitate dropping the restriction to planar motion, and extension of earlier results to the spatial problem.

Dynamical Channels in the Spatial Problem. From now on, we focus on the spatial problem. We will put more emphasis on numerical computations, especially issues concerning halo orbit missions. In this chapter we present a detailed portrait of the phase space geometry near L_1 and L_2. In Chapter 9, we start from this portrait to consider the dynamical channels connecting the realms in the spatial problem connected by bottlenecks about L_1 and L_2. We extend the tube dynamics to the spatial problem and present an example multi-moon orbiter using the higher dimensional tube methodology. Tube dynamics in the spatial problem involves considering the invariant manifolds of a three-sphere of bound orbits, similar to the role of the cylindrical invariant manifolds of a Lyapunov periodic orbit, a one-sphere, in the planar problem.

Outline of the Chapter. In this chapter, we first discuss some history of missions using CR3BP solutions, particularly the halo orbits. We then introduce the CR3BP and discuss the halo orbits, three-dimensional periodic orbits about the libration points, and discuss how to compute them. The computation of halo orbits follows standard nonlinear trajectory computation algorithms based on differential correction. Due to the sensitivity of the problem, an accurate first guess is essential, since the halo orbit is actually an unstable orbit (albeit with a fairly long time constant in the Sun-Earth system, about 180 days). This first guess is provided by a high order analytic expansion using the Lindstedt–Poincaré method. In order to gain familiarity with the Lindstedt–Poincaré method, we first consider a textbook example, the Duffing equation. We then use the Lindstedt–Poincaré method to find a third-order approximation of a halo orbit, and use this approximation as a first guess in a differential correction scheme. Finally, we discuss briefly the orbit structure near L_1 and L_2, laying the foundation for later chapters.

6.2 Some History of Halo Orbits for Space Missions

We review some of the history of halo orbits and key missions for which the methods described in this chapter were developed.

Discovery of Halo Orbits. The points of equilibrium along the line joining two massive bodies, in particular L_1 and L_2, have been known for over two centuries (Euler [1767] and Lagrange [1772]). Their existence was

known to the pioneers of spaceflight in the mid-twentieth century, but no application of them was proposed until 1950. At that time, futurist Arthur C. Clarke suggested that the translunar libration point, L_2, would be an ideal place for relaying TV and radio broadcasts from Earth to colonists on the far side of the Moon (Clarke [1950]). A decade and a half later, Robert Farquhar discovered trajectories around L_2 in which a communication satellite could be placed allowing continuous link between the Earth and the Moon's far side (Farquhar [1966]). He named such a trajectory a "halo orbit" as it appeared from the Earth to be a halo encircling the Moon. In space, a halo orbit looks like the edges of a potato chip,

During the late 1960s and early 1970s, NASA's Apollo program was underway, and many people were working on aspects of manned lunar operations. Farquhar published numerous papers extensively covering topics related to halo orbits, libration point satellite operations in general, and their applications within the Earth-Moon system and beyond (Farquhar [1968, 1969, 1972]).[1] Other authors conducted comprehensive studies of halo orbit families for the Earth-Moon system (Breakwell and Brown [1979]), and for all other two mass gravity fields (Howell and Breakwell [1984], and Howell [1984]).

International Sun-Earth Explorer/International Cometary Explorer. In 1972, the International Sun-Earth Explorer (ISEE) Program was estabilished, a joint project of NASA and the European Space Agency (ESA) which was to involve three spacecraft. One of these, ISEE-3, was launched into a halo orbit around the Sun-Earth L_1 point in 1978, allowing it to collect data on solar wind conditions upstream from Earth. ISEE-3 accomplished many scientific goals. After the primary mission was completed, ISEE-3 went on to accomplish other goals, including a flight through the geomagnetic tail and a comet flyby, by utilizing the interesting dynamics in the Sun-Earth-Moon system (Farquhar [2001]). The mission was subsequently renamed the International Cometary Explorer (ICE).

Barcelona Group and SOHO. In the mid-1980s, a team in Barcelona were the first to study the invariant manifolds of halo orbits and apply them to the design of a space mission, in particular the Solar and Heliospheric Observatory (SOHO) mission (see references Simó, Gómez, Llibre, and Martinez [1986] and Simó, Gómez, Llibre, Martinez, and Rodríguez [1987]). SOHO did not require the delicate controls provided by this theory, so the actual mission was flown using the classical methods developed at NASA by Farquhar and coworkers (see, for example, Farquhar and Dunham [1981]; Farquhar, Muhonen, Newman, and Heuberger [1980]; Farquhar, Muhonen, and Richardson [1977]).

[1] Unfortunately, NASA scheduled no further moon landings after 1972, and the agency's desire for a communication link with the Moon's far side disappeared (see Farquhar [2001]).

Other Missions. Lissajous and halo type trajectories around the collinear libration points have been considered in the trajectory design of many space missions, including the Genesis mission and others mentioned in the supplement in §1.2. In the future, the European Space Agency is considering two missions to L_2 in the Sun-Earth system, FIRST and PLANK. Furthermore, for NASA's Next Generation Space Telescope, the follow-on to the Hubble Telescope, an L_2 libration point orbit is has been considered. This interest in libration point orbits justifies the study of the dynamics around an extended neighborhood of these points in order that more complex missions can be envisaged.

6.3 A Few Basic Facts About the CR3BP

Energy Surface and Hill's Region. The level surfaces of the Hamiltonian energy (2.3.13) , $E = e =$ constant, which are also ***energy surfaces***, are invariant 5-dimensional manifolds. Let $\mathcal{M}$ be that energy surface, i.e.,

$$\mathcal{M}(\mu, e) = \{(x, y, z, \dot{x}, \dot{y}, \dot{z}) \mid E(x, y, z, \dot{x}, \dot{y}, \dot{z}) = e\}.$$

The projection of this surface onto position space is called a ***Hill's region***

$$M(\mu, e) = \{(x, y, z) \mid \bar{U}(x, y, z) \leq e\}.$$

The boundary of $M(\mu, e)$ is the ***zero velocity surface***. The intersection of this surface with the xy-plane, i.e., $z = 0$, is the zero velocity curve in the xy-plane. The spacecraft can move only within this region, the white region in Figure 2.4.2.

Our main concern in the rest of the book is the behavior of the orbits of equations (2.3.8) whose energy is just above that of L_2; that is, $e > E_2$, where the value of the energy integral at the point L_i is denoted by E_i. For this case, the three-dimensional Hill's region contains a "neck" about L_1 and L_2, as shown in Figure 6.3.1. Thus, orbits with an energy just above that of L_2 are energetically permitted to make a transit through the two neck regions from the ***interior realm*** (inside the Earth's orbit) to the ***exterior realm*** (outside the Earth's orbit) passing through the ***Earth (capture) realm***. Moreover, as shown later in this chapter, the computation of halo orbits also requires the opening of the neck about L_1 or L_2.

Equilibria of the CR3BP. By applying the same techniques used in Chapter 2, it can be shown that the system (2.3.8) has five equilibrium points, all of which are in the xy-plane: three collinear points on the x-axis, called L_1, L_2, L_3 (see Figure 6.3.1) and two equilateral points called L_4 and L_5. These equilibrium points are critical points of the (effective potential) function $\bar{U}$.

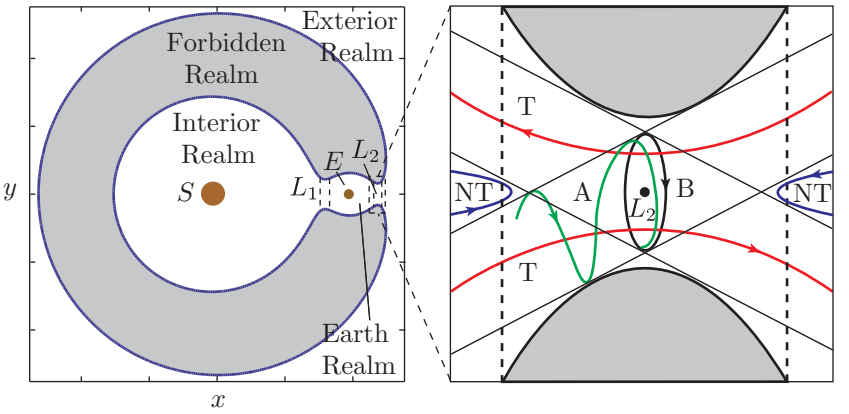


FIGURE 6.3.1. (a) A two-dimensional slice ($z = 0$) of the three-dimensional Hill's region (schematic, the region in white), which contains three large realms connected by neck regions about L_1 and L_2. (b) The flow in the L_2 neck region is shown projected on the xy-plane, showing a bounded orbit around L_2 (labeled B), an asymptotic orbit winding onto the bounded orbit (A), two transit orbits (T) and two non-transit orbits (NT), shown in schematic. A similar figure holds for the region around L_1.

Computation of the values of the abscissas of the collinear points requires the solution of $\frac{d}{dx}\bar{U}(x,0,0) = 0$, and was covered in §2.5. The distance from $L_i, i = 1, 2$ to the smaller primary is given by γ_i, the unique positive solution of (2.5.2).

Equations of Motion near L_1 or L_2. The equations of motion for a satellite moving in the vicinity of L_1 or L_2 can be obtained by translating the origin to the location of L_1 or L_2. The change of coordinates is given by

$$\begin{aligned}\bar{x} &= \frac{x - 1 + \mu \pm \gamma}{\gamma},\\ \bar{y} &= \frac{y}{\gamma},\\ \bar{z} &= \frac{z}{\gamma},\end{aligned}$$

where the upper sign is for L_1 and the lower sign is for L_2. In this new ***normalized coordinate system***, the variables $\bar{x}, \bar{y}, \bar{z}$ are scaled so that the distance between L_1 or L_2 and the small primary is 1. This scaled system was introduced in Richardson [1980a] in order to have good numerical properties for the series expansion below. For simplicity of notation, x, y, z will still be used to denote the variables in the normalized coordinate system for the rest of the chapter. For the Sun-Earth L_1, $\gamma_1 = 1.001090475 \times 10^{-2}$ and one unit of normalized distance $= 1.497610041 \times 10^6$ km (where we use

a slightly different Sun-Earth distance than the one given in Table 2.2.1, $L = 1.495978714 \times 10^8$ km).

Expansion of the Nonlinear Equations. Some computational advantages can be obtained if the equations are developed using Legendre polynomials P_n. In order to expand the nonlinear terms, $\frac{1-\mu}{r_1} + \frac{\mu}{r_2}$, the following formula can be used,

$$\frac{1}{\sqrt{(x-A)^2+(y-B)^2+(z-C)^2}} = \frac{1}{D}\sum_{n=0}^{\infty}\left(\frac{\rho}{D}\right)^n P_n\left(\frac{Ax+By+Cz}{D\rho}\right),$$

where $D^2 = A^2 + B^2 + C^2$ and $\rho^2 = x^2 + y^2 + z^2$. After some calculations, the equations of motion can be written as

$$\begin{aligned}
\ddot{x} - 2\dot{y} - (1+2c_2)x &= \frac{\partial}{\partial x}\sum_{n\geq 3} c_n \rho^n P_n\left(\frac{x}{\rho}\right)\\
\ddot{y} + 2\dot{x} + (c_2-1)y &= \frac{\partial}{\partial y}\sum_{n\geq 3} c_n \rho^n P_n\left(\frac{x}{\rho}\right) \qquad (6.3.1)\\
\ddot{z} + c_2 z &= \frac{\partial}{\partial z}\sum_{n\geq 3} c_n \rho^n P_n\left(\frac{x}{\rho}\right)
\end{aligned}$$

where the left-hand side contains the linear terms and the right-hand side contains the nonlinear ones. The coefficients $c_n(\mu)$ are given by

$$c_n = \frac{1}{\gamma^3}\left((\pm)^n\mu + (-1)^n\frac{(1-\mu)\gamma^{n+1}}{(1\mp\gamma)^{n+1}}\right).$$

where the upper sign is for L_1 and the lower one for L_2. Notice that c_2 is the same as $\bar{\mu}$ (cf. (2.6.2)) from the planar problem discussion of Chapter 2. For the Sun-Earth L_1, $c_2 = 4.0610735668, c_3 = 3.0200105081$, and $c_4 = 3.0305378797$.

These expressions for the equations of motion are in an ideal form from the point of view that all orders of the nonlinear expansion can be developed recursively using the well-known Legendre polynomial relationships (see Whittaker [1927]). For instance, if we define

$$T_n(x,y,z) = \rho^n P_n\left(\frac{x}{\rho}\right),$$

then T_n is a homogeneous polynomial of degree n that satisfies the recurrence

$$T_n = \frac{2n-1}{n}xT_{n-1} - \frac{n-1}{n}(x^2+y^2+z^2)T_{n-2},$$

starting with $T_0 = 1$ and $T_1 = x$. This is particularly useful if the successive approximation solution procedure is carried to high orders via algebraic manipulation software programs.

6.4 Overview of Halo Orbit Computations

As mentioned previously, the absence of a general solution for the CR3BP has motivated a number of researchers to develop approximate analytical solutions. Linear analysis suggests the existence of periodic (and quasi-periodic) orbits near the collinear libration points. Higher order approximations can provide further insight into the general nature of such solutions. They can be used as a first guess for the numerical computation of the desired orbits. The halo orbit used by the first halo orbit mission ISEE-3 was designed using this methodology. See Farquhar and Kamel [1973] and Richardson [1980b].

To date, a complete description and computation of the libration orbits around an extended neighborhood of the collinear libration points has been obtained. The main objects found are planar and vertical families of Lyapunov periodic orbits; three-dimensional quasi-periodic Lissajous orbits; periodic halo orbits; and quasi-halo orbits. For more details, see Howell and Pernicka [1988], Gómez, Jorba, Masdemont, and Simó [1991a], and Jorba and Masdemont [1999].

Here, we will follow the presentation of Richardson [1980b] which gives a succinct summary on how the particular ISEE-3 halo orbit was constructed using analytical and numerical methods. The third-order analytical solution was obtained by an application of successive approximations using the Lindstedt–Poincaré method. The ISEE-3 mission orbit was then produced from standard differential correction procedures using the third-order solution as the "first guess".

Periodic Solutions of the Linearized Equations. The existence of periodic solutions to the nonlinear equations (6.3.1) can be deduced by considering the linear part of the equations:

$$\begin{aligned} \ddot{x} - 2\dot{y} - (1 + 2c_2)x &= 0 \\ \ddot{y} + 2\dot{x} + (c_2 - 1)y &= 0 \\ \ddot{z} + c_2 z &= 0 \end{aligned} \tag{6.4.1}$$

It is clear that the z-axis solution, obtained by putting $x = 0$ and $y = 0$, is simple-harmonic since $c_2 > 0$ and does not depend on x and y. However, the motion in the xy-plane is coupled. As we know from §2.6 and §2.7, the solution of the characteristic equation of (6.4.1) has two real and two imaginary roots $(\pm\lambda, \pm i\omega_p)$. Hence, the linear behavior near the collinear libration point is of the type saddle×center×center with eigenvalues $(\pm\lambda, \pm i\omega_p, \pm i\omega_v)$ given by

$$\lambda^2 = \frac{c_2 - 2 + \sqrt{9c_2^2 - 8c_2}}{2}, \qquad \omega_p^2 = \frac{2 - c_2 + \sqrt{9c_2^2 - 8c_2}}{2}, \qquad \omega_v^2 = c_2.$$

Since the two real roots are opposite in sign, arbitrarily chosen initial conditions will give rise, in general, to unbounded solutions as time in-

creases. If, however, the initial conditions are restricted so that only the non-divergent mode is allowed, the xy-solution will be bounded. In this case, the linearized equations have solutions of the form

$$\begin{aligned} x &= -A_x \cos(\omega_p t + \phi) \\ y &= \kappa A_x \sin(\omega_p t + \phi) \\ z &= A_z \sin(\omega_v t + \psi) \end{aligned} \tag{6.4.2}$$

with

$$\kappa = \frac{\omega_p^2 + 1 + 2c_2}{2\omega_p}. \tag{6.4.3}$$

The linearized motion will become quasi-periodic if the in-plane and out-of-plane frequencies are such that their ratio is irrational. The projections of the motion onto the various coordinate planes produce Lissajous-type trajectories, as shown in Figure 6.4.1. Notice that κ is the same as $-\tau$ (cf. (2.7.2)) from Chapter 2.

Halo Orbits in the Linearized Approximation. For the Sun-Earth L_1, $\omega_p = 2.086453455$ and $\omega_v = 2.0152105515$. Their difference is quite small. Motion similar on a short time scale to periodic (halo) motion is obtained if the amplitudes of the in-plane and out-of-plane motions are of sufficient magnitude so that the nonlinear contributions to the system produce eigenfrequencies that are equal. For the ISEE-3 halo, shown in Figure 6.4.2, $A_z = 110{,}000$ km, $A_x = 206{,}000$ km, and $A_y = \kappa A_x = 665{,}000$ km (with $\kappa = 3.2292680962$).

Equations (6.4.2) with $\omega_p = \omega_v$ will form the first approximation for the third-order periodic solution using the Lindstedt–Poincaré procedure of successive approximations. Details will be given below in §6.5 and §6.6. Here, we will highlight some of the interesting results.

Amplitude Constraint Relationship. For halo orbits, the amplitudes A_x and A_z are constrained by a certain non-linear algebraic relationship found as result of the applicaiton of the perturbation method:

$$l_1 A_x^2 + l_2 A_z^2 + \Delta = 0. \tag{6.4.4}$$

For Sun-Earth L_1 halo orbits, $l_1 = -15.9650314$, $l_2 = 1.740900800$, and $\Delta = 0.29221444425$.

Hence, any halo orbit can be characterized completely by specifying a particular out-of-ecliptic plane amplitude A_z of the solution to the linearized equations of motion. Both the analytical and numerical developments employ this identifying scheme. The ISEE-3 spacecraft was targeted to a halo orbit corresponding to an A_z amplitude of 110,000 km.

From the expression (6.4.4), one can find the minimum permissible value for A_x in order to have a halo orbit ($A_z > 0$). In the case of a periodic solution about L_1 in the Sun-Earth system, the value of A_x is about 14%

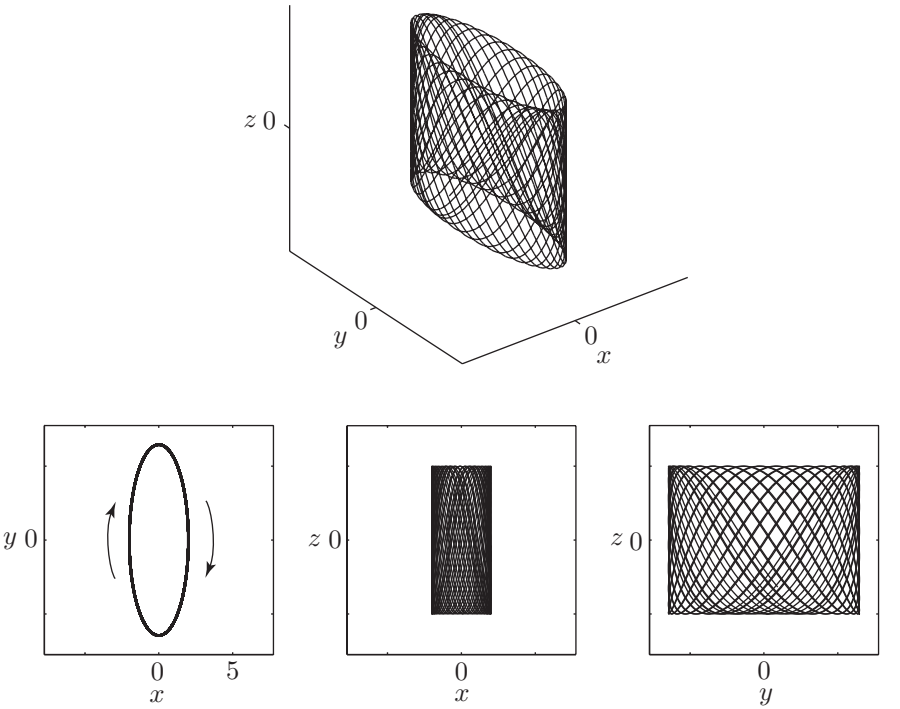


FIGURE 6.4.1. (top) The general motion around a libration point is shown. (bottom) The same motion projected onto the various coordinate planes produce Lissajous-type trajectories. The trajectory is shown around the Sun-Earth L_1 point as the origin. Arrows indicate the clockwise direction of motion in the xy projection. The values of A_x and A_z are not important and are chosen only to illustrate the typical Lissajous patterns.

of the normalized distance. This amounts to a minimum amplitude of approximately 200,000 km.

Phase-Angle Relationship. For halo orbits, the phases ϕ and ψ are related to each other in a linear fashion:

$$\psi - \phi = m\pi/2, \qquad m = 1, 3. \tag{6.4.5}$$

It will be shown later that when A_x is greater than a certain minimum value, the third-order solution bifurcates. This bifurcation manifests itself through the phase-angle constraint relation and two solution branches are obtained according to whether $m = 1$ or $m = 3$. The two branches are viewed as mirror reflections of each other about the xy-plane. For $m = 1$,

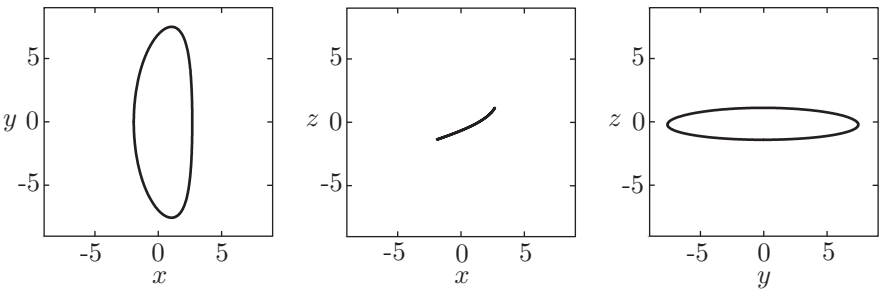


FIGURE 6.4.2. The halo orbit used by the ISEE-3 mission. From left to right, the xy, xz, and yz projections are shown. The Sun-Earth L_1 point is the origin, with coordinate values given in increments of 10^5 km.

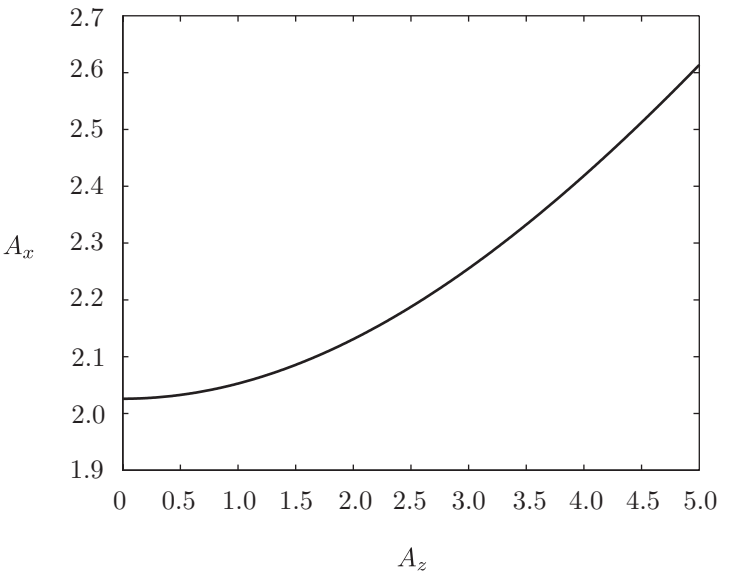


FIGURE 6.4.3. The amplitude constraint relationship for Sun-Earth L_1 halo orbits. The coordinate values are given in increments of 10^5 km.

A_z is positive, and we have the northern (class I) halo whose maximum out-of-plane component is above the xy-plane ($z > 0$). For $m = 3$, A_z is negative, and we have the southern (class II) halo whose maximum out-of-plane component is below the xy-plane ($z < 0$). Northern and southern halo orbits with the same A_z amplitude are mirror images across the xy-plane. In Figure 6.4.4, a northern and southern halo orbit about the Sun-Earth L_2 are shown.

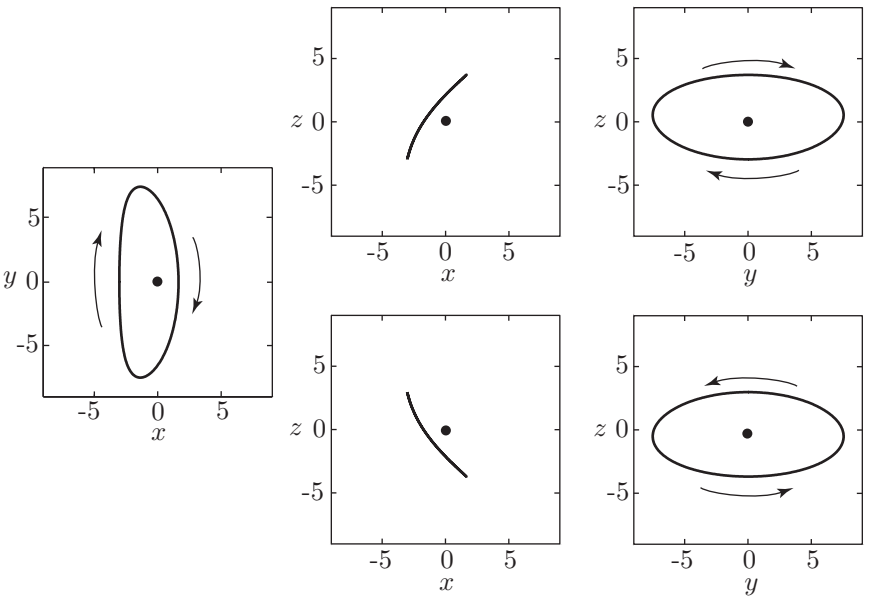


FIGURE 6.4.4. A northern halo (top) and southern halo (bottom) about the Sun-Earth L_2. From left to right, the xy, xz, and yz projections are shown. The xy projections are the same. Arrows indicate the direction of motion in the xy and yz projections. The Sun-Earth L_2 point is the dot at the origin. Coordinate values are given in increments of 10^5 km. $|A_z|$=330,000 km for both halos.

6.5 Periodic Orbits and the Lindstedt–Poincaré Method

Periodic orbits are of fundamental importance in the study of dynamical systems in general, and Hamiltonian systems in particular. They are also important destinations for spacecraft missions, e.g. halo orbits. In order to compute families of periodic orbits, one often uses analytical approximations. The ***Lindstedt–Poincaré method*** is one such method which we will use.

To introduce and illustrate the Lindstedt–Poincaré method, let us study the Duffing equation,

$$\ddot{q} + q + \epsilon q^3 = 0, \tag{6.5.1}$$

where ϵ is a small parameter.

Finding a Periodic Solution for the Duffing Equation. For $\epsilon = 0$, (6.5.1) has a periodic solution

$$q = a \cos t,$$

for the initial conditions $q(0) = a, \dot{q}(0) = 0$.

For $\epsilon \neq 0$, we look for a periodic solution. A naive approach would be to try to find a solution of the following form,

$$q = \sum_{n=0}^{\infty} \epsilon^n q_n(t) = q_0(t) + \epsilon q_1(t) + \epsilon^2 q_2(t) + \cdots, \tag{6.5.2}$$

with the initial conditions $q(0) = a, \dot{q}(0) = 0$. This implies that $q_0(0) = a, \dot{q}_0(0) = 0$ and $q_n(0) = 0, \dot{q}_n(0) = 0$ for $n \geq 1$.

By substituting the solution (6.5.2) into equation (6.5.1) and equating terms having same power of ϵ, one gets a system of recursive differential equations:

$$\begin{aligned} \ddot{q}_0 + q_0 &= 0, \\ \ddot{q}_1 + q_1 &= -q_0^3, \\ \ddot{q}_2 + q_2 &= -3q_0^2 q_1, \end{aligned}$$

and so on. Clearly, the solution for q_0 in the first equation with initial conditions $q_0(0) = a, \dot{q}_0(0) = 0$ is

$$q_0 = a \cos t.$$

Substituting this expression for q_0 into the second equation gives

$$\ddot{q}_1 + q_1 = -q_0^3 = -a^3 \cos^3 t = -\frac{1}{4} a^3 (\cos 3t + 3 \cos t).$$

The solution for q_1 with initial conditions $q_1(0) = 0, \dot{q}_1(0) = 0$ is

$$q_1 = -\frac{3}{8} a^3 t \sin t + \frac{1}{32} a^3 (\cos 3t - \cos t).$$

The term $t \sin t$ is called a ***secular term*** and such terms grow in amplitude as t grows. Due to the presence of secular terms, a naive method such as an expansion of the solution in a power series in ϵ *does not work*. The Lindstedt–Poincaré method is designed to get around this difficulty.

The Lindstedt–Poincaré Method. The main idea behind the Lindstedt–Poincaré method builds on the observation that *non-linearity* alters the *frequency* of the linearized system. This change of frequency, from 1 to $\nu(\epsilon)$, has to be taken into consideration, and it can be done efficiently by introducing a new independent variable,

$$\tau = \nu(\epsilon) t,$$

so that for the rescaled Duffing equation,

$$q'' + \nu^{-2}(q + \epsilon q^3) = 0, \tag{6.5.3}$$

with τ as the independent variable, the desired periodic solution will still be $2\pi-$periodic. Expand ν^{-1} as follow

$$\nu^{-1} = 1 + \epsilon\nu_1 + \epsilon^2\nu_2 + \cdots .$$

Correspondingly, expand the desired periodic solution in a power series in ϵ:

$$q = \sum_{n=0}^{\infty} \epsilon^n q_n(\tau) = q_0(\tau) + \epsilon q_1(\tau) + \epsilon^2 q_2(\tau) + \cdots . \tag{6.5.4}$$

Substitute the power series expansion of q (6.5.4) into the rescaled Duffing equation (6.5.3) and equate terms with the same power of ϵ. This gives equations for successive approximations:

$$\begin{aligned}
&q_0'' + q_0 = 0,\\
&q_1'' + q_1 = -q_0^3 - 2\nu_1 q_0,\\
&q_2'' + q_2 = -3q_0^2 q_1 - 2\nu_1(q_1 + q_0^3) - (\nu_1^2 + 2\nu_2)q_0,
\end{aligned}$$

and so on.

Now we come to a key point: potential secular terms can be avoided by imposing suitable conditions on ν_n. Again, the solution for q_0 in the first equation with the initial conditions $q_0(0) = a, \dot{q}(0) = 0$ is

$$q_0 = a\cos\tau.$$

By substituting $q_0 = a\cos\tau$ into the second equation, we get

$$\begin{aligned}
q_1'' + q_1 &= -a^3\cos^3\tau - 2\nu_1 a\cos\tau,\\
&= -\frac{1}{4}a^3\cos 3\tau - \left(\frac{3}{4}a^2 + 2\nu_1\right)a\cos\tau.
\end{aligned}$$

In previous naive method, we had $\nu_1 \equiv 0$ and a secular term appeared because the $\cos\tau$ term has the same frequency as the oscillator $q_1'' + q_1$ on the left-hand-side of the equation. However, if we set $\nu_1 = -3a^2/8$, we can get rid of the $\cos\tau$ term so that no secular term appears. Then the solution for the second equation is given by,

$$q_1 = -\frac{1}{32}a^3\left(\cos\tau - \cos 3\tau\right),$$

for the initial conditions $q_1(0) = 0, \dot{q}(0) = 0$.

Therefore, to first-order in ϵ, we have a periodic solution

$$\begin{aligned}
q &= a\cos\tau - \frac{1}{32}\epsilon a^3\left(\cos\tau - \cos 3\tau\right) + O(\epsilon^2),\\
&= a\cos\nu t - \frac{1}{32}\epsilon a^3\left(\cos\nu t - \cos 3\nu t\right) + O(\epsilon^2),
\end{aligned}$$

with

$$\begin{aligned}\nu &= \left\{1 - \epsilon\nu_1 - \frac{1}{2}\epsilon^2(2\nu_2 - \nu_1^2) + \cdots\right\},\\ &= 1 + \frac{3}{8}\epsilon a^2 + O(\epsilon^2).\end{aligned}$$

In summary, the Lindstedt–Poincaré method involves successive adjustments of frequencies to avoid secular terms and allows one to obtain approximate periodic solutions.

6.6 Third-Order Richardson Expansion

Recall that the CR3BP equations can be developed using Legendre polynomials P_n, as in (6.3.1). A third-order approximation was used in Richardson [1980a]:

$$\begin{aligned}\ddot{x} - 2\dot{y} - (1 + 2c_2)x &= \frac{3}{2}c_3(2x^2 - y^2 - z^2)\\ &\quad + 2c_4x(2x^2 - 3y^2 - 3z^2) + O(4),\\ \ddot{y} + 2\dot{x} + (c_2 - 1)y &= -3c_3xy - \frac{3}{2}c_4y(4x^2 - y^2 - z^2) + O(4),\\ \ddot{z} + c_2z &= -3c_3xz - \frac{3}{2}c_4z(4x^2 - y^2 - z^2) + O(4).\end{aligned}$$

Construction of Periodic Solutions. As explained earlier, halo-type periodic solutions are obtained by assuming that the amplitudes A_x and A_z of the linearized solution (6.4.2) are large enough so that the nonlinear contributions to the system produce eigenfrequencies which are equal ($\omega_p = \omega_v$). With this assumption, the linearized equations can be rewritten as

$$\begin{aligned}\ddot{x} - 2\dot{y} - (1 + 2c_2)x &= 0,\\ \ddot{y} + 2\dot{x} + (c_2 - 1)y &= 0,\\ \ddot{z} + \omega_p^2 z &= 0,\end{aligned} \tag{6.6.1}$$

where ω_p^2 has replaced the coefficient c_2 in the last equation of (6.4.1). These linear equations have a periodic solution with frequency ω_p

$$\begin{aligned}x &= -A_x\cos(\omega_p t + \phi),\\ y &= \kappa A_x\sin(\omega_p t + \phi),\\ z &= A_z\sin(\omega_p t + \psi),\end{aligned} \tag{6.6.2}$$

which can be used as the seed for constructing successive approximations.

Correction Term. By forcing the linearized z equation (6.4.1) to the form (6.6.1), it becomes necessary to rewrite the left-hand side of the z equation of (6.6.1) and to introduce a correction term $\Delta = \omega_p^2 - c_2 = \omega_p^2 - \omega_v^2$ on the right-hand side when higher-order approximations are constructed. The new third-order z equation then becomes,

$$\ddot{z} + \omega_p^2 z = -3c_3 xz - \frac{3}{2}c_4 z(4x^2 - y^2 - z^2) + \Delta z + O(4).$$

It can be shown that for motion about L_1 in the Sun-Earth system that Δ should obey the order-of-magnitude relation $\Delta = 0.29221 = O(A_z^2)$. Accordingly, the Δz contribution to the solution will first appear in the expressions for the third-order corrections.

Lindstedt–Poincaré Method. Richardson [1980a] developed a third-order periodic solution using a Lindstedt–Poincaré type of successive approximations (cf. §6.5). To help remove secular terms, a new independent variable τ and a frequency connection ν are introduced via,

$$\tau = \nu t.$$

Here,

$$\nu = 1 + \sum_{n \geq 1} \nu_n, \qquad \nu_n < 1.$$

The ν_n are assumed to be $O(A_z^n)$ and are chosen to remove secular terms as they appear during the course of the development of the successive approximations solution. Notice that $A_z \ll 1$ in normalized units and it plays the role of ϵ.

The equations of motion are then written in terms of new independent variable τ

$$\begin{aligned}
\nu^2 x'' - 2\nu y' - (1 + 2c_2)x &= \frac{3}{2}c_3(2x^2 - y^2 - z^2) \\
&\quad + 2c_4 x(2x^2 - 3y^2 - 3z^2) + O(4), \\
\nu^2 y'' + 2\nu x' + (c_2 - 1)y &= -3c_3 xy \\
&\quad - \frac{3}{2}c_4 y(4x^2 - y^2 - z^2) + O(4), \\
\nu^2 z'' + \omega_p^2 z &= -3c_3 xz \\
&\quad - \frac{3}{2}c_4 z(4x^2 - y^2 - z^2) + \Delta z + O(4).
\end{aligned}$$

To find the third-order successive approximation solution is a lengthy process. Here are some highlights.

The generating solution used is the linearized solution (6.6.2) with t replaced by τ. Most of the secular terms can be removed by proper specification of the ν_n. It is found that

$$\nu_1 = 0, \qquad \nu_2 = s_1 A_x^2 + s_2 A_z^2, \tag{6.6.3}$$

which gives the frequency $\omega_p\nu$ and the period ($T = 2\pi/\omega_p\nu$) of a halo orbit.

However, not all of the secular terms can be removed in this manner. Additionally, it becomes necessary to specify amplitude and phase angle constraint relationships, as mentioned earlier. These expressions are

$$l_1 A_x^2 + l_2 A_z^2 + \Delta = 0, \tag{6.6.4}$$

$$\psi - \phi = m\pi/2, \qquad m = 1, 3. \tag{6.6.5}$$

Halo Orbits in the Third-Order Approximation. The third-order solution in Richardson [1980a] is given by:

$$\begin{aligned}
x &= a_{21}A_x^2 + a_{22}A_z^2 - A_x\cos\tau_1 \\
&\quad +(a_{23}A_x^2 - a_{24}A_z^2)\cos 2\tau_1 + (a_{31}A_x^3 - a_{32}A_xA_z^2)\cos 3\tau_1, \\
y &= \kappa A_x \sin\tau_1 \\
&\quad +(b_{21}A_x^2 - b_{22}A_z^2)\sin 2\tau_1 + (b_{31}A_x^3 - b_{32}A_xA_z^2)\sin 3\tau_1, \\
z &= \delta_m A_z \cos\tau_1 \\
&\quad +\delta_m d_{21}A_xA_z(\cos 2\tau_1 - 3) + \delta_m(d_{32}A_zA_x^2 - d_{31}A_z^3)\cos 3\tau_1.
\end{aligned}$$

where $\tau_1 = \omega_p\tau + \phi$ and $\delta_m = 2 - m, m = 1, 3$. Two solution branches are obtained according to whether $m = 1$ or $m = 3$.

Values of Constants. The constants s_i, l_i, and a_{ij}, b_{ij} and d_{ij} are complicated expressions involving c_2, c_3 and c_4, which we reproduce here.

$$\begin{aligned}
a_{21} &= \frac{3c_3(\kappa^2 - 2)}{4(1 + 2c_2)}, \\
a_{22} &= \frac{3c_3}{4(1 + 2c_2)}, \\
a_{23} &= -\frac{3c_3\omega_p}{4\kappa d_1}\left(3\kappa^3\omega_p - 6\kappa(\kappa - \omega_p) + 4\right), \\
a_{24} &= -\frac{3c_3\omega_p}{4\kappa d_1}(2 + 3\kappa\omega_p), \\
b_{21} &= -\frac{3c_3\omega_p}{2d_1}(3\kappa\omega_p - 4), \\
b_{22} &= -\frac{3c_3\omega_p}{d_1}, \\
d_{21} &= -\frac{c_3}{2\omega_p^2}, \\
a_{31} &= -\frac{9\omega_p}{4d_2}\left(4c_3(\kappa a_{23} - b_{21}) + \kappa c_4(4 + \kappa^2)\right) \\
&\quad + \frac{9\omega_p^2 + 1 - c_2}{2d_2}\left(3c_3(2a_{23} - \kappa b_{21}) + c_4(2 + 3\kappa^2)\right), \\
a_{32} &= -\frac{9\omega_p}{4d_2}\left(4c(3\kappa a_{24} - b_{22}) + \kappa c_4\right)
\end{aligned}$$

$$
\begin{aligned}
&- \frac{3}{2d_2}\left(9\omega_p^2+1-c_2\right)\left(c_3(\kappa b_{22}+d_{21}-2a_{24})-c_4\right),\\
b_{31} =& \frac{3}{8d_2}8\omega_p\left(3c_3(\kappa b_{21}-2a_{23})-c_4(2+3\kappa^2)\right)\\
&+\frac{3}{8d_2}\left((9\omega_p^2+1+2c_2)(4c_3(\kappa a_{23}-b_{21})+\kappa c_4(4+\kappa^2)\right),\\
b_{32} =& \frac{9\omega_p}{d_2}\left(c_3(\kappa b_{22}+d_{21}-2a_{24})-c_4\right)\\
&+\frac{3(9\omega_p^2+1+2c_2)}{8d_2}\left(4c_3(\kappa a_{24}-b_{22})+\kappa c_4\right),\\
d_{31} =& \frac{3}{64\omega_p^2}\left(4c_3a_{24}+c_4\right),\\
d_{32} =& \frac{3}{64\omega_p^2}\left(4c_3(a_{23}-d_{21})+c_4(4+\kappa^2)\right),\\
s_1 =& \left(2\omega_p(\omega_p(1+\kappa^2)-2\kappa)\right)^{-1}\\
&\times\left(\frac{3}{2}c_3(2a_{21}(\kappa^2-2)-a_{23}(\kappa^2+2)-2\kappa b_{21})-\frac{3}{8}c_4(3\kappa^4-8\kappa^2+8)\right),\\
s_2 =& \left(2\omega_p(\omega_p(1+\kappa^2)-2\kappa)\right)^{-1}\\
&\times\left(\frac{3}{2}c_3(2a_{22}(\kappa^2-2)+a_{24}(\kappa^2+2)+2\kappa b_{22}+5d_{21})+\frac{3}{8}c_4(12-\kappa^2)\right),\\
l_1 =& -\frac{3}{2}c_3\left(2a_{21}+a_{23}+5d_{21}\right)-\frac{3}{8}c_4\left(12-\kappa^2\right)+2\omega_p^2 s_1,\\
l_2 =& \frac{3}{2}c_3\left(a_{24}-2a_{22}\right)+\frac{9}{8}c_4+2\omega_p^2 s_2,
\end{aligned}
$$

where κ is given by (6.4.3) and d_1, d_2 are

$$
\begin{aligned}
d_1 =& \frac{3\omega_p^2}{\kappa}\left(\kappa(6\omega_p^2-1)-2\omega_p\right),\\
d_2 =& \frac{8\omega_p^2}{\kappa}\left(\kappa(11\omega_p^2-1)-2\omega_p\right).
\end{aligned}
$$

For the Sun-Earth L_1, $s_1 = -8.246608317 \times 10^{-1}$, $s_2 = 1.210985938 \times 10^{-1}$, $l_1 = -1.596560314 \times 10^{1}$ and $l_2 = 1.740900800$.

Halo Orbit Phase-Angle Relationship. A bifurcation manifests through the phase-angle relationship:

- For $m = 1$, $\delta_m > 0$. Northern halo.
- For $m = 3$, $\delta_m < 0$. Southern halo.

Clearly, northern and southern halos are mirror images of each other across the xy-plane. See, for instance, Figure 6.4.4.

Halo Orbit Amplitude Constraint Relationship. Since there exists an amplitude constraint relationship (6.6.4), the minimum value for A_x to have a halo orbit (i.e., for $A_z > 0$) is $\sqrt{|\Delta/l_1|}$, which is about 200,000 km. Moreover, a halo orbit can be characterized completely by its z-amplitude A_z. For the ISEE-3 halo with A_z =110,000 km, its x-amplitude was,

$$A_x = 206,000 \text{ km},$$

which corresponds to a y-amplitude of,

$$A_y = \kappa A_x = 665,000 \text{ km}.$$

Halo Orbit Period Amplitude Relationship. The halo orbit period $T = 2\pi/\omega_p\nu$ can be computed as a function of A_z via the amplitude constraint relationship (6.6.4) and the frequency connection ν (6.6.3). This relationship is shown in Figure 6.6.1. The ISEE-3 halo had a period of 177.73 days.

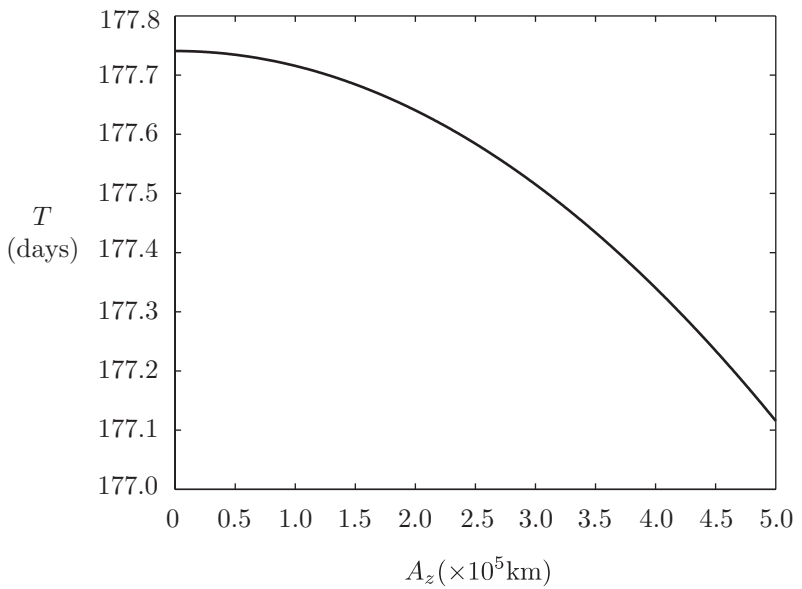


FIGURE 6.6.1. The halo orbit period (T) vs. z amplitude (A_z) relationship for Sun-Earth L_1 halo orbits. The halo orbits have periods of a little less than half an Earth year.

6.7 Numerical Computation of Halo Orbits

While third-order approximations provide much qualitative insight, they are insufficient for serious study of accurate motion near L_1 or L_2. Analytical approximations must be combined with numerical techniques to generate a halo orbit accurate enough for mission design. This problem is well suited to a differential correction process discussed in §4.2, using the Richardson approximation as the first guess.

State Transition Matrix for CR3BP. The state transition matrix along a reference orbit for the CR3BP,

$$\delta\bar{x}(t) = \Phi(t, t_0)\delta\bar{x}(t_0),$$

can be computed numerically by integrating simultaneously the following 42 ODEs:

$$\dot{\bar{x}} = f(\bar{x}), \tag{6.7.1}$$
$$\dot{\Phi}(t, t_0) = Df(\bar{x})\Phi(t, t_0), \tag{6.7.2}$$

with initial conditions:

$$\bar{x}(t_0) = \bar{x}_0,$$
$$\Phi(t_0, t_0) = I_6.$$

Here, equations (6.7.1) are CR3BP equations (2.3.8) written in the first-order form. The Jacobian matrix $Df(\bar{x})$ is

$$Df(\bar{x}) = \begin{pmatrix} 0 & I_3 \\ -\mathcal{U} & 2\Omega \end{pmatrix},$$

where matrix Ω can be written as,

$$\Omega = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Matrix $\mathcal{U}$ has the form,

$$\mathcal{U} = \begin{pmatrix} \bar{U}_{xx} & \bar{U}_{xy} & \bar{U}_{xz} \\ \bar{U}_{yx} & \bar{U}_{yy} & \bar{U}_{yz} \\ \bar{U}_{zx} & \bar{U}_{zy} & \bar{U}_{zz} \end{pmatrix},$$

and is evalutated along the reference solution. Matrix I_3 is the 3×3 identity matrix.

Numerical Computation of Halo Orbits. To begin the procedure of using differential correction for finding a periodic solution, both an initial state and a final state must be defined. Symmetries of halo orbits can be

used to aid this step. Recall that halo orbits are symmetric about the xz-plane ($y = 0$), and they intersect this plane perpendicularly ($\dot{x} = \dot{z} = 0$). Thus, the initial state vector takes the form,

$$\bar{x}_0 = (x_0, 0, z_0, 0, \dot{y}_0, 0)^T.$$

Obtain a first guess for the elements of this state vector from the third-order approximations. The equations of motion are then integrated until the trajectory crosses the xz-plane. To produce a periodic orbit, a perpendicular crossing is needed, and so the desired final (target) state vector needs to have the form,

$$\bar{x}_f = (x_f, 0, z_f, 0, \dot{y}_f, 0)^T.$$

Of course, the first crossing of the xz-plane occurs at a time equal to one half of the period of the orbit. It is likely, however, that actual values for $\dot{x}_f$ and $\dot{z}_f$ are not zero at the first crossing. The three non-zero initial conditions (x_0, z_0, and $\dot{y}_0$) can be manipulated in an attempt to drive these final velocities ($\dot{x}_f$ and $\dot{z}_f$) to zero.

Recall that differential correction uses the state transition matrix to change initial conditions,

$$\delta\bar{x}_f = \Phi(t_f, t_0)\delta\bar{x}_0.$$

The change $\delta\bar{x}_0$ can be determined by the difference between actual and desired final states ($\delta\bar{x}_f = \bar{x}_f^d - \bar{x}_f$). Three initial states ($\delta x_0, \delta z_0, \delta\dot{y}_0$) are available to target two final states ($\delta\dot{x}_f, \delta\dot{z}_f$).

Rather than invert a two by three matrix subject to a minimum norm to find the appropriate changes for x_0, z_0, and $\dot{y}_0$, it is more convenient to constrain one of the initial state variables, e.g., $\delta z_0 = 0$. This results in a 2×2 matrix which is easily inverted to find variations δx_0 and $\delta\dot{y}_0$.

Similarly, the revised initial conditions $\bar{x}_0 + \delta\bar{x}_0$ are used to begin a second iteration. This process is continued until $\dot{x}_f = \dot{z}_f = 0$ (within some acceptable tolerance). Usually, convergence to a solution is achieved within a few (≤ 5) iterations.

6.8 Libration Orbits near Collinear Points

In last decade or so, members of the Barcelona group such as Gómez, Jorba, Masdemont and Simó have done an extensive study of the dynamics of libration orbits near the collinear points. To obtain an accurate description, two different but complementary approaches have been used. On the one hand, an effective computation of the center manifold of the collinear points has been devised to give an overall qualitative description of the dynamics of these libration orbits. On the other hand, a Lindstedt–Poincaré method

of higher order expansions has been used to compute various types of libration orbits, such as halo, Lissajous, and "quasi-halo" orbits (discussed later in this section). In this and the next section, we will describe briefly these two approaches and their results. For more details, the reader can consult Chapter 9 or the original papers, such as Gómez, Jorba, Masdemont, and Simó [1991a], Gómez, Masdemont, and Simó [1997, 1998], and Jorba and Masdemont [1999].

Linear Behavior and Expansion of the Hamiltonian. Recall from equation (2.3.11) that the Hamiltonian is given by

$$\begin{aligned} H(x,y,z,p_x,p_y,p_z) &= p_x\dot{x} + p_y\dot{y} + p_z\dot{z} - L \\ &= \frac{1}{2}\left((p_x+y)^2 + (p_y-x)^2 + p_z^2\right) + \bar{U}(x,y,z), \end{aligned} \tag{6.8.1}$$

where p_x, p_y and p_z are the conjugate momenta to the variables (x, y, z).

Using similar techniques as on the Lagrangian side, the linearization of the Hamiltonian around $L_{1,2}$ shows that the local behavior near these points are of the type saddle×center×center. So, using a real, linear, and symplectic change of coordinates, it is easy to cast the second-order part of the preceding Hamiltonian

$$H_2 = \frac{1}{2}\left(p_x^2 + p_y^2\right) + yp_x - xp_y - c_2x^2 + \frac{c_2}{2}y^2 + \frac{1}{2}p_z^2 + \frac{c_2}{2}z^2,$$

into its real normal form,

$$H_2 = \lambda x p_x + \frac{\omega_p}{2}(y^2 + p_y^2) + \frac{\omega_v}{2}(z^2 + p_z^2). \tag{6.8.2}$$

Note that for simplicity, we have kept in equation (6.8.2) the same notation for the new variables even after a coordinate change.

For the following computations it is convenient to "diagonalize" the second-order terms. This is done by introducing the complex change of coordinates

$$\begin{aligned} \begin{pmatrix} q_2 \\ p_2 \end{pmatrix} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} y \\ p_y \end{pmatrix}, \\ \begin{pmatrix} q_3 \\ p_3 \end{pmatrix} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} z \\ p_z \end{pmatrix}, \end{aligned} \tag{6.8.3}$$

and renaming $q_1 = x$ and $p_1 = p_x$, the second-order part of the Hamiltonian becomes

$$H_2 = \lambda q_1 p_1 + \sqrt{-1}\omega_p q_2 p_2 + \sqrt{-1}\omega_v q_3 p_3 \tag{6.8.4}$$

From now on we will use the following notation. If $x = (x_1, \ldots, x_n)$ is a vector of complex numbers and $k = (k_1, \ldots, k_n)$ is an integer vector,

we denote by x^k the term $x_1^{k_1} \cdots x_n^{k_n}$ (in this context we define 0^0 as 1). Moreover, we define $|k|$ as $\sum_j |k_j|$.

In order to have all possible orbits in the center manifold, let us expand the initial Hamiltonian H (6.8.1) using the coordinates that give us H_2 as in (6.8.4). Then the expanded Hamiltonian takes the form

$$\begin{aligned} H(q,p) &= H_2(q,p) + \sum_{n\geq 3} H_n(q,p), \\ &= H_2(q,p) + \sum_{n\geq 3} \sum_{|k|=n} h_k q_1^{k_1} p_1^{k_2} q_2^{k_3} p_2^{k_4} q_3^{k_5} p_3^{k_6}, \end{aligned} \tag{6.8.5}$$

where H_2 is given in (6.8.4) and H_n denotes an homogeneous polynomial of degree n.

Reduction to the Center Manifold. The next step is to perform canonical transformations to (6.8.5) to obtain

$$H(q,p) = H_2(q,p) + \sum_{n\geq 3} \sum_{|k|=n, k_1=k_2} h_k q_1^{k_1} p_1^{k_2} q_2^{k_3} p_2^{k_4} q_3^{k_5} p_3^{k_6}.$$

The purpose is to "kill" all the monomials with $k_1 \neq k_2$. Of course, this is done up to a finite order N. Call $I = q_1 p_1$, and the Hamiltonian looks like

$$H = H_N(I, q_2, p_2, q_3, p_3) + R_{N+1}(q,p).$$

Now, if we drop the remainder R_{N+1} which is very small near $L_{1,2}$, we see that I is a first integral. Setting $I = 0$, we skip the hyperbolic part and reduce the Hamiltonian to its center manifold.

Finally, using the inverse change of variables of (6.8.3), the truncated Hamiltonian H_N can be expanded in real form and we obtain

$$H_N(0, q_2, p_2, q_3, p_3) = H_2 + \sum_{n=3}^{N} H_n(0, q_2, p_2, q_3, p_3),$$

where explicitly,

$$H_2 = H_2(0, q_2, p_2, q_3, p_3) = \frac{\omega_p}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega_v}{2}\left(q_3^2 + p_3^2\right).$$

For convenience, the new variables are called again q, p.

All the computations have been implemented by writing specific symbolic manipulators in the Fortran or C programming languages that can do all the procedures up to an arbitrary order. For practical purposes, and in order to have an acceptable balance between precision and computing time, the numerical scheme has been implemented up to order $N = 15$.

Libration Orbits around Collinear Points. As shown in the previous section, orbits in the center manifold can be obtained by setting $q_1 = p_1 = 0$ in the initial conditions, $(q_1, p_1, q_2, p_2, q_3, p_3)$. Moreover, if we consider only orbits of same energy, three free variables remain. Looking at the orbits when they cross a surface of section, all the libration orbits with a selected energy can be obtained from only two variables in the initial conditions. For instance, the initial conditions can be chosen selecting arbitrary values for q_2 and p_2, by setting $q_3 = 0$ (the surface of section) and finally computing p_3 in order to be in the selected level of Hamiltonian energy. The propagation of these initial conditions, looking when and where they crosses the surface of section again and again, gives what is called the images of the Poincaré map in the variables (q_2, p_2) on $q_3 = 0$.

However, we want to see the orbits in CR3BP coordinates. For this purpose we can take the initial conditions as before, but we look at the Poincaré map of the orbit when it crosses and re-crosses the plane $z = 0$ in the CR3BP coordinates. The orbits are then clearly seen by plotting their (x, y)-CR3BP coordinates on the section. We note that due to the linear part of the CR3BP equations of motion around the collinear equilibrium points, $z = 0$ is a surface of section for all the libration orbits in a neighborhood of the equilibrium point, except for the planar ones (the ones having $z = \dot{z} = 0$) which are contained in the $z = 0$ plane.

This is the procedure that we have used to obtain Figure (6.8.1), where the libration orbits around L_1 and L_2 are displayed for some values of the energy of the CR3BP giving qualitative different pictures.

We note that each level of energy we have a bounded region in the Poincaré section. The boundary of the plot is a planar Lyapunov orbit of the selected energy contained in the surface of section. It is the only one orbit contained in the xy-plane and it is essentially related to the frequency ω_p of H_2. The fixed point, in the central part of the figures, corresponds to an almost vertical periodic orbit (see Figure 6.8.2), essentially related to the frequency ω_v of H_2. Surrounding the central fixed point, we have invariant curves corresponding to Lissajous orbits. The motion in this region is essentially quasi-periodic (except for very small chaotic zones that cannot be seen in the pictures).

Depending on the value of the energy integral, there appear two fixed points close to the boundaries of the plot. A fixed point means again a periodic orbit, and in this case, they are the well known halo orbits of class I (northern) and class II (southern). Surrounding the fixed points corresponding to the halo orbits, we have again invariant curves related to quasi-periodic motions. They are Lissajous orbits that we call ***quasi-halo orbits***.

Finally, in the transition zone from central Lissajous to quasi-halo orbits there is an homoclinic connection of the planar Lyapunov orbit. We note that the homoclinic trajectory that goes out from the orbit and the one that goes in, do not generally coincide; they intersect with a very small

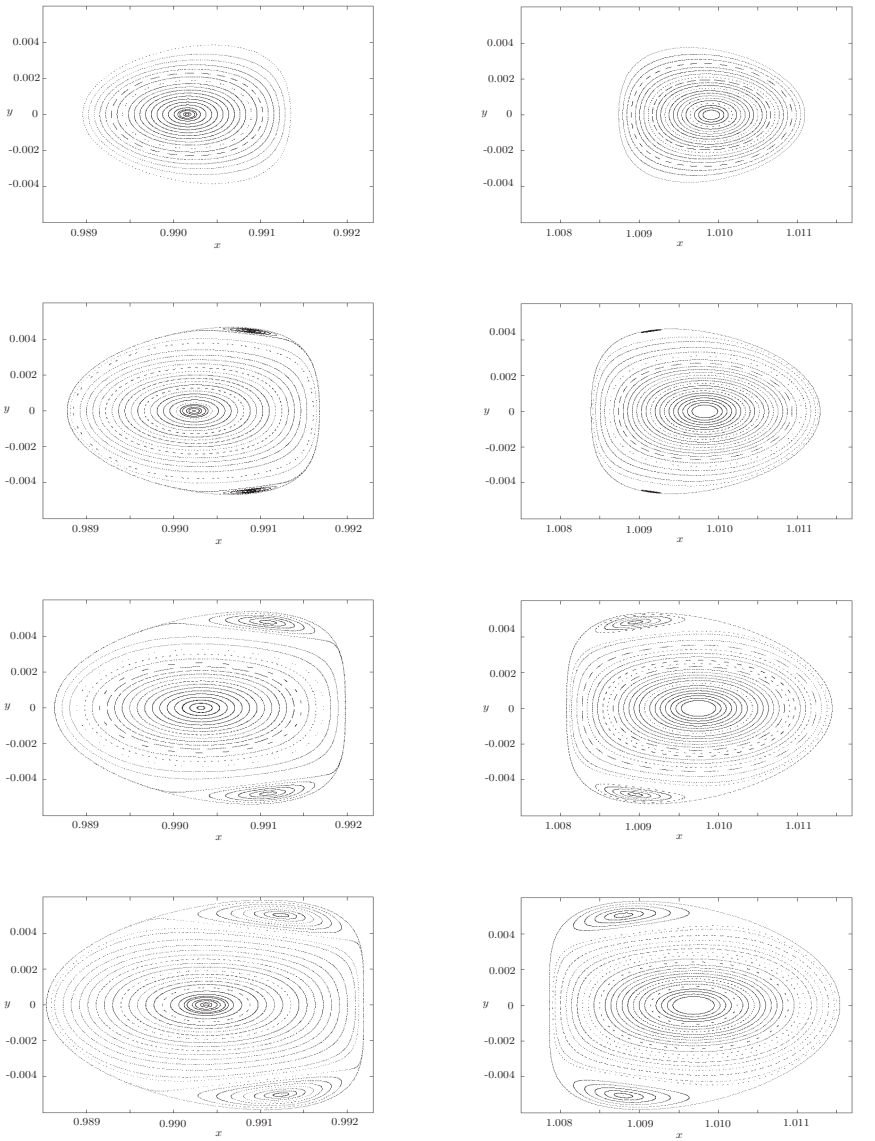


FIGURE 6.8.1. Poincaré maps of the orbits in the central manifold of L_1 (four left figures) and L_2 (four right figures) for the following four increasing values of the energy from top to bottom: -1.500425, -1.50041322952164, -1.50040145756682, and -1.50039257918817. The mass parameter used is that of the Sun-Earth system, $\mu = 3.040423398444176 \times 10^{-6}$. Normalized coordinates are used in the figures.

angle. This phenomenon is known as splitting of separatrices. We also note in this case, that the planar Lyapunov orbit is unstable even in the central manifold.

In the next section we will show how the planar and vertical Lyapunov, Lissajous, halo and quasi-halo family of orbits can be computed using Lindstedt–Poincaré procedures and ad hoc algebraic manipulators. In this way one obtains their expansions in CR3BP coordinates suitable to be used in a friendly way. In Figure 6.8.2 a sample of all these orbits is shown.

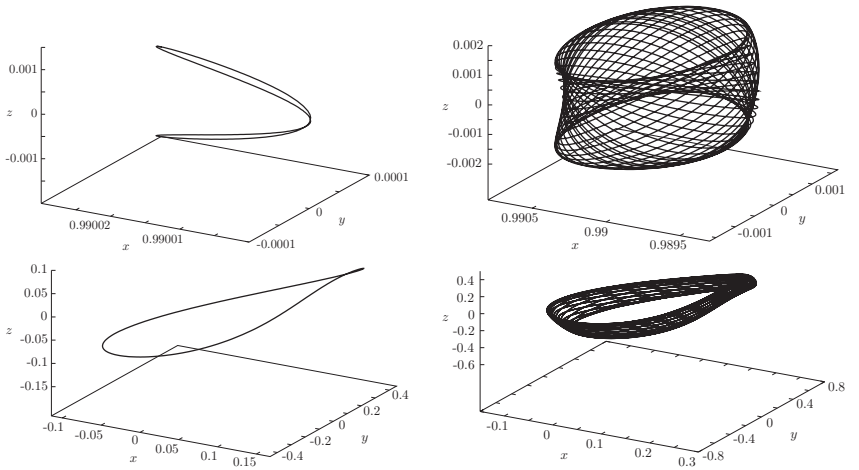


FIGURE 6.8.2. 3D representation of several types of orbits around L_1. Upper left: vertical periodic orbit with $\alpha = 0.0$ and $\beta = 0.1$. Upper right: Lissajous orbit with $\alpha = 0.05$ and $\beta = 0.15$. Lower left: halo orbit with $\beta = 0.1$. Lower right: quasi-halo orbit with $\beta = 0.2$ and $\gamma = 0.067$. The upper plots are in CR3BP coordinates. The lower plots are in L_1 normalized coordinates.

6.9 Computation of Libration Orbits in the Center Manifold

Computation of Lissajous Orbits. We will start with the computation of the Lissajous trajectories (two dimensional tori) and the halo orbits (one dimensional tori) using the Lindstedt–Poincaré method. The exposition in this section follows the unpublished notes "State of the Art" by Gómez and Masdemont. If we consider the linear part of the system (6.3.1), its solution can be written as

$$x(t) = -\alpha \cos(\omega_p t + \phi_1),$$

$$\begin{aligned} y(t) &= \kappa\alpha \sin(\omega_p t + \phi_1), \\ z(t) &= \beta \cos(\omega_v t + \phi_2), \end{aligned} \tag{6.9.1}$$

where $\omega_p = \sqrt{\frac{2-c_2+\sqrt{9c_2^2-8c_2}}{2}}$, $\omega_v = \sqrt{c_2}$ and $\kappa = \frac{\omega_p^2+1+2c_2}{2\omega_p}$. The parameters α and β are the in-plane and out-of-plane amplitudes of the orbit and ϕ_1, ϕ_2 are the phases. These linear solutions are already Lissajous trajectories. When we consider the nonlinear terms of the equations, we look for formal series solutions in powers of the amplitudes α and β of the type

$$\begin{aligned} x &= \sum_{i,j=1}^{\infty} \left(\sum_{|k|\le i, |m|\le j} a_{ijkm} \cos(k\theta_1 + m\theta_2) \right) \alpha^i \beta^j, \\ y &= \sum_{i,j=1}^{\infty} \left(\sum_{|k|\le i, |m|\le j} b_{ijkm} \sin(k\theta_1 + m\theta_2) \right) \alpha^i \beta^j, \\ z &= \sum_{i,j=1}^{\infty} \left(\sum_{|k|\le i, |m|\le j} c_{ijkm} \cos(k\theta_1 + m\theta_2) \right) \alpha^i \beta^j, \end{aligned} \tag{6.9.2}$$

where $\theta_1 = \omega t + \phi_1$ and $\theta_2 = \nu t + \phi_1$. Due to the presence of nonlinear terms, the frequencies ω and ν cannot be kept equal to ω_p and ω_v, and they must be expanded in powers of the amplitudes

$$\begin{aligned} \omega &= \sum_{i,j=0}^{\infty} \omega_{ij} \alpha^i \beta^j = \omega_p + \sum_{i,j=1}^{\infty} \omega_{ij} \alpha^i \beta^j, \\ \nu &= \sum_{i,j=0}^{\infty} \nu_{ij} \alpha^i \beta^j = \omega_v + \sum_{i,j=1}^{\infty} \nu_{ij} \alpha^i \beta^j. \end{aligned}$$

The goal is to compute the coefficients a_{ijkm}, b_{ijkm}, c_{ijkm}, ω_{ij}, and ν_{ij} recursively up to a finite order $N = i + j$. First, identifying the coefficients of the general solution (6.9.2) with the ones obtained from the solution of the linear part (6.9.1), we see that only non zero values are $a_{1010} = 1$, $b_{1010} = \kappa$, $c_{1010} = 1$, $\omega_{00} = \omega_p$ and $\nu_{00} = \omega_v$. Then, inserting the linear solution (6.9.1) in (6.3.1), we get a reminder for each equation, which is a series in α and β beginning with terms of order $i + j = 2$. In order to get the coefficients of order 2, the known order 2 terms must be equated to the unknown order 2 terms of the left hand side. The general step is similar. It assumes that the solution has been computed up to a certain order $n - 1$. Then it is substituted in the right-hand side of (6.3.1) producing terms of order n in α and β. These known order n terms must be equated with the unknown terms of order n of the left-hand side.

This procedure can be implemented up to high orders and the results obtained are satisfactory. In this way we get, close to the equilibrium point,

a big set of KAM tori. In fact, between these tori there are very narrow stochastic zones (because the resonances are dense). Hence we will have divergence everywhere. However, small divisors will show up only at high orders (except the one due to the 1:1 resonance), because at the origin ω_p/ω_v is close to 29/28. The high order resonances have a very small stochastic zone and the effect is only seen after a long time interval. It is not difficult to find "practical" convergence regions in the (α, β)-plane by just comparing the formal series solution evaluated at some epoch, and the solution obtained using numerical integration of the full equations of motion. In Figure 6.9.1, we show these convergence regions by displaying the values of α and β for which the difference between the analytical expansions (truncated at different orders) and the result of the numerical integration after π time units is lower than 10^{-6} non-dimensional CR3BP units.

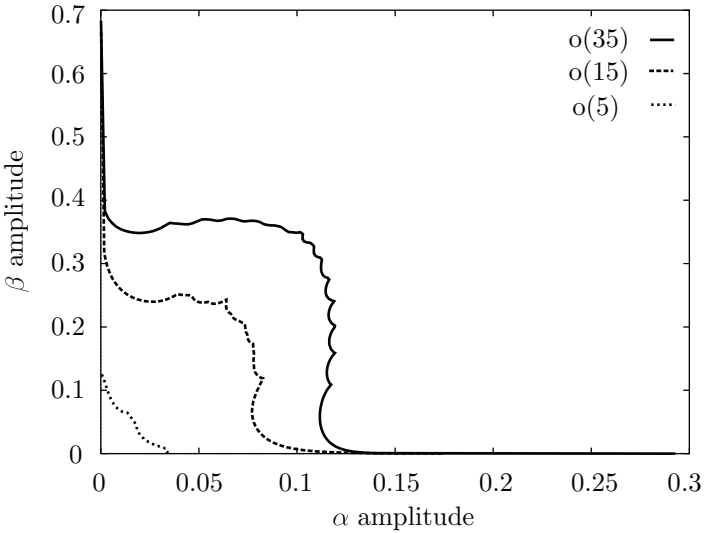


FIGURE 6.9.1. Domain of practical convergence of the Lindstedt–Poincaré expansion of the Lissajous orbits around L_1 in the Earth-Sun system. The different lines correspond to orders 5, 15 and 35 for the expansions.

Computation of Halo Orbits to High Order. As discussed earlier, halo orbits are periodic orbits which bifurcate from the planar Lyapunov periodic orbits (Lissajous orbits with $\beta = 0$) when the in plane and out of plane frequencies are equal. This is a 1:1 resonance that appears as a consequence of the nonlinear terms of the equations. Of course we have to look for these one-dimensional invariant tori as series expansion with a single frequency. These solutions do not appear as a solution of the linearized equations. In order to apply the Lindstedt–Poincaré procedure we modify the equations of motion (6.3.1) as before by adding the product of

the factors Δ and z to the third equation, where Δ is a frequency type series,

$$\Delta = \sum_{i,j=0}^{\infty} f_{ij}\alpha^i\beta^j,$$

that must verify the condition $\Delta = 0$. As before, we start looking for the (non trivial) librating solutions with one frequency of the linearized equations, which is

$$\begin{aligned} x(t) &= \alpha\cos(\omega_p t + \phi_1),\\ y(t) &= \kappa\alpha\sin(\omega_p t + \phi_1),\\ z(t) &= \beta\cos(\omega_p t + \phi_2). \end{aligned} \tag{6.9.3}$$

We note that after this first step, halo orbits are determined up to order 1, and $\Delta = 0$ is read as $f_{00} = 0$. Halo orbits depend only on one frequency or one amplitude since they are one-dimensional invariant tori, so we do not have two independent amplitudes α and β. The relation between α and β is contained in the condition $\Delta = 0$ which defines implicitly $\alpha = \alpha(\beta)$.

When we consider the full equations, we look for formal expansions in powers of the amplitudes α and β of the type

$$\begin{aligned} x(t) &= \sum_{i,j=1}^{\infty}\left(\sum_{|k|\le i+j} a_{ijk}\cos(k\theta)\right)\alpha^i\beta^j,\\ y(t) &= \sum_{i,j=1}^{\infty}\left(\sum_{|k|\le i+j} b_{ijk}\sin(k\theta)\right)\alpha^i\beta^j,\\ z(t) &= \sum_{i,j=1}^{\infty}\left(\sum_{|k|\le i+j} d_{ijk}\cos(k\theta)\right)\alpha^i\beta^j, \end{aligned}$$

where $\theta = \omega t + \phi$ and, as in the case of two-dimensional invariant tori, the frequency ω must be expanded as $\omega = \sum_{i,j=0}^{\infty}\omega_{ij}\alpha^i\beta^j$. The procedure for the computation of the unknown coefficients a_{ijk}, b_{ijk}, d_{ijk}, ω_{ij} and f_{ij} is similar to the one described for the Lissajous trajectories.

As halo orbits constitute a one parameter family of periodic orbits, they can be identified by a single parameter. This can be the normalized z-amplitude, β, of the orbit, that is, the amplitude of the first harmonic in the expansion of the z variable, in the normalized coordinates previously introduced. For small amplitudes they can also be uniquely identified by the frequency ω.

Computation of Quasi-Halo Orbits. Quasi-halo orbits are quasi-periodic orbits on two-dimensional tori around a halo orbit. They depend on two basic frequencies. Given a halo orbit of frequency ω, the series expansions

for the coordinates of the quasi-halo orbits will be of the form

$$x(t) = \sum_{i=1}^{\infty} \left(\sum_{|k|<i,|m|<i} a_i^{km} \cos\left(k(\omega t + \phi_1) + m(\nu t + \phi_2)\right) \right) \gamma^i,$$

$$y(t) = \sum_{i=1}^{\infty} \left(\sum_{|k|<i,|m|<i} b_i^{km} \sin\left(k(\omega t + \phi_1) + m(\nu t + \phi_2)\right) \right) \gamma^i,$$

$$z(t) = \sum_{i=1}^{\infty} \left(\sum_{|k|<i,|m|<i} c_i^{km} \cos\left(k(\omega t + \phi_1) + m(\nu t + \phi_2)\right) \right) \gamma^i,$$

These expansions depend on two frequencies, ω and ν, and one amplitude, γ. This amplitude is related to the size of the torus around the "base" halo orbit which is taken as the backbone. The frequency ν is the second natural frequency of the torus, and it is close to the normal frequency around the base halo orbit.

In order to apply the Lindstedt–Poincaré method to compute the quasi-halo orbits, it is convenient to perform a change of variables which transforms the halo orbit to an equilibrium point of the equations of motion. Then, orbits librating around the equilibrium point in the new coordinates correspond to orbits librating around the halo orbit in the original ones. The details of the procedure can be found in Gómez, Masdemont, and Simó [1997, 1998].

7

Invariant Manifolds and End-to-End Transfer

7.1 Introduction

A number of old as well as recent missions have been designed to take advantage of the scientific interest in the region of space near libration points in the Sun-Earth system (see §6.2 and §5.4). In support of missions that include increasingly complex trajectories and incorporate libration point orbits, more efficient techniques and new philosophies for design must be considered. In this chapter, the Genesis mission (see supplement in §1.2) is used to demonstrate the usefulness of dynamical systems theory in trajectory design. Of particular significance is the use of invariant manifolds to produce a transfer from the Earth to a Sun-Earth libration point orbit and to return the spacecraft back to the Earth with a day side re-entry. Furthermore, the methodology used to meet launch and return constraints will also be presented. See Barden [1994], Howell, Barden, and Lo [1997], and Lo, Williams, Bollman, Han, Hahn, Bell, Hirst, Corwin, Hong, Howell, Barden, and Wilson [2001] and references therein.

Invariant Manifolds. Transfer in the three body problem has not been investigated extensively until the last few decades. Since there are few analytical tools available for this problem, researchers must depend on numerical approaches which were not readily available until the 1960s. Apparently, the first in-depth analysis of transfers between an orbit about a

primary and a libration point was in D'Amario [1973] which studied the fuel-optimal transfer between collinear libration points and circular orbits around either the Earth or the Moon. In 1980, Farquhar et al. published their post-mission analysis results on the transfer trajectory between the Earth and the nominal L_1 halo orbit used by ISEE-3 (Farquhar, Muhonen, Newman, and Heuberger [1980]). However, Gómez, Jorba, Masdemont, and Simó [1991b] were the first to use invariant manifold theory to aid in determining transfers to a libration point orbit from an Earth parking orbit. By building on the results of Gómez et al., Howell et al. combined invariant manifold methods with differential correction schemes to create a design tool useful for generating transfers from an Earth parking orbit to a halo orbit in the Sun-Earth system. In this chapter, we will follow the approach of Howell, Barden, Wilson, and Lo [1997].

The optimal Genesis trajectory was found by computing and understanding the characteristics of the invariant manifolds associated with its halo orbit. The transfer trajectory was constructed using the stable manifold of the L_1 halo orbit which approaches the Earth (in reverse time), some pieces of which are shown in Figure 7.1.1. Following the completion of the halo orbit phase of the mission, the vehicle automatically leaves the libration point region with no departure maneuver, as the return trajectory was constructed to be on the halo orbit's unstable manifold, Figure 7.1.2.

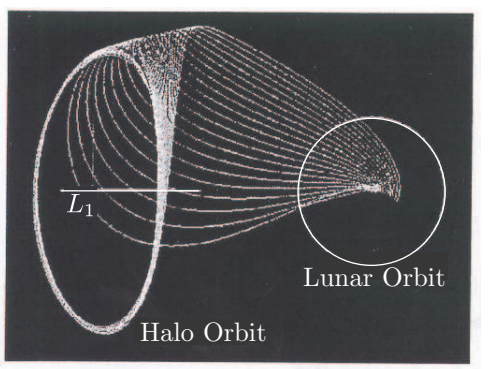


FIGURE 7.1.1. Stable manifold used to design the Genesis transfer orbit.

Since the return was designed as a segment of the unstable manifold associated with the halo orbit, the orbit naturally unwinds from the halo shape. The requirement to return to the Utah Test and Training Range for a mid-air helicopter recovery required that the vehicle return during daylight hours. A trajectory from the unstable manifold that provides a dayside Earth return was found; however, it requires that the vehicle first enter the L_2 region before returning to the Earth. This free transfer exploits the heteroclinic behavior of the L_1 and L_2 regions. This transport mechanism,

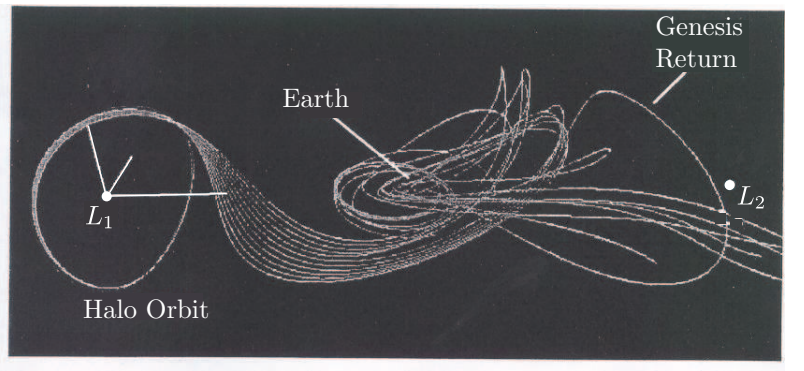


FIGURE 7.1.2. Unstable manifold used to design the Genesis return orbit.

which was discussed in Chapter 3 is what governs the temporary planetary capture of comets and provides the free return of the Genesis mission. The region of space that can be reached with the unstable manifolds of the orbit is vast; however, the segment of the manifold that provides an acceptable return to Earth is extremely small as illustrated in Figure 7.1.2. The Earth return requires very precise targetting of a very narrow corridor of phase space.

Outline of the Chapter. While the study of the periodic orbit in the context of the restricted three-body problem has a long history, only recently have significant advances been made in issues concerning transfer trajectories to these types of orbits using invariant manifolds.

In §7.2, we discuss how to approximate stable and unstable manifolds of a halo orbit, based on the stability analysis of a periodic orbit discussed in §4.3. These approximations are then used as the first step in a numerical search for transfer trajectories from an Earth parking orbit to a halo orbit in the vicinity of the L_1 or L_2 libration point. In §7.3 and §7.4, we will discuss how to incorporate a simple differential correction scheme, the Halo Orbit Insertion (HOI) corrector, into the design process (see Howell, Mains, and Barden [1994]). The numerical results in their work include direct and indirect transfers from the Earth to halo orbits near L_1 and L_2.

In §7.5, the differential corrector used to find the end-to-end Genesis trajectory will be discussed. Frequently, a simple differential corrector such as the HOI corrector may not work. The trajectory may be too "stiff," i.e., the initial condtions are too sensitive or there are too many constraints at the end point than there are parameters at the starting point. In these cases, one can break the trajectory into segments by adding ΔVs and using a *two-level differential correction scheme* to get a solution that meets specified launch and return constraints. Differential correction is schematically illustrated in Figure 7.1.3 and will be discussed in detail at the end of the

chapter.

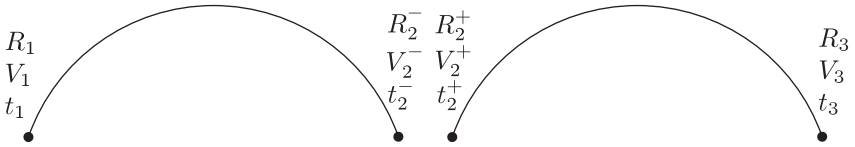


FIGURE 7.1.3. Use differential correction to patch various pieces of solution arcs together into a trajectory that meets specified launch and return constraints. The points represent position (R) and velocities (V) at certain times (t_i) and the arcs connecting them are solutions of the equations of motion. Differential correction is a multiple shooting method which patches together solution arcs.

7.2 Computation of Invariant Manifolds Associated to a Halo Orbit

The computation of the stable and unstable manifolds associated with a particular halo orbit can be accomplished numerically. The procedure is based on the monodromy matrix discussed in §4.3 which is the state transition matrix after one period of a periodic orbit (the halo orbit, in this case). For any point $\bar{x}_0$ along the halo orbit (a fixed point of a Poincaré map), the mondromy matrix at $\bar{x}_0$ serves essentially as the linearization of the Poincaré map near the fixed point. As with the linearization of any discrete mapping of a fixed point, the characteristics of the local geometry of the phase space can be determined from the eigenvalues and eigenvectors of the monodromy matrix. The eigenvectors can then be used in approximating the local invariant manifolds near $\bar{x}_0$.

Eigenvalues and Eigenvectors of the Monodromy Matrix in CR3BP. Recall that the state transition matrix $\Phi(t)$ along a halo orbit can be obtained numerically by integrating the variational equations on the halo orbit at the same time that the halo orbit is computed (see §6.7). After that, the eigenvalues of the monodromy matrix can also be computed numerically. For halo orbits, two of the eigenvalues are 1 and the remaining four eigenvalues include one real pair and one complex pair. In summary,

$$\lambda_1 > 1, \qquad \lambda_2 = \frac{1}{\lambda_1}, \qquad \lambda_3 = \lambda_4 = 1, \qquad , \lambda_5 = \bar{\lambda}_6, \qquad |\lambda_5| = 1.$$

where λ_5 and λ_6 are complex conjugates, and λ_1 and λ_2 are real. The determinant of any matrix is equal to the product of its eigenvalues and the determinant of the monodromy matrix is equal to one by Liouville's

theorem. Thus, since λ_1 and λ_2 are real, the moduli of λ_5 and λ_6 must be one.

1. The first pair (λ_1, λ_2), with $\lambda_1 \cdot \lambda_2 = 1$ and $\lambda_1 \gg 1$, is associated with the unstable characteristic of the small and medium sized halo orbits. (For example, $\lambda_1 \approx 1500$ for halo orbits around the Sun-Earth L_1 and L_2.) They describe dynamics which are hyperbolic in nature (see Figure 4.4.2). The value λ_1 is the *dominant* eigenvalue. The eigenvector associated to λ_1 (we call it $\mathbf{e}_1(0)$) gives the most expanding direction. Using the state transition matrix, we can obtain the image of $\mathbf{e}_1(0)$ under the variational flow: $\mathbf{e}_1(\tau) = \Phi(\tau, 0) \cdot \mathbf{e}_1(0)$. At each point of the halo orbit, the vector $\mathbf{e}_1(\tau)$, together with the vector tangent to the orbit, span a plane. This plane is tangent to the local unstable manifold. By forward continuation of points under the flow, we can produce the full unstable manifold.

   Using the symmetry (3.3.1) we obtain the stable manifold and, in particular, the vector $\mathbf{e}_2(\tau) = \Phi(\tau, 0)\mathbf{e}_2(0)$. The eigenvalue of $\mathbf{e}_2(0)$ is λ_2.

2. The second pair, $(\lambda_3, \lambda_4) = (1, 1)$, is associated with the neutral variables (i.e., the non-unstable modes). However, there is only one (right) eigenvector of M with eigenvalue equal to one. This vector $\mathbf{e}_3(t)$ is the tangent vector to the halo orbit. We call it $\mathbf{e}_3(0)$ at time $t = 0$ and $\mathbf{e}_3(t)$ in general.

   The other eigenvalue, $\lambda_4 = 1$, is associated with variations of the energy, or any equivalent variable. For small size halo orbits, a suitable variable associated with this eigenvalue is a parameter of the family of halo orbits, such as the z-amplitude or the period.

   Along the orbit, these two vectors generate a family of planes which are sent from one into another by the variational flow. The monodromy matrix restricted to this plane has the following Jordan canonical form

$$\begin{pmatrix} 1 & \epsilon \\ 0 & 1 \end{pmatrix}.$$

   The fact that ϵ is not zero is due to the variation of the period when the orbit changes across the family.

3. The third pair is associated to the complex eigenvalues $\lambda_5 = \bar{\lambda}_6$ of modulus one. The monodromy matrix, restricted to the plane spanned by the corresponding pair of eigenvectors associated to λ_5, λ_6 has the following form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

i.e., it is a rotation. This behavior is related to the existence of quasiperiodic orbits around the halo orbit, the quasi-halos discussed in §6.8. For more information, see Gómez, Masdemont, and Simó [1998].

Use Eigenvectors of the Monodromy Matrix as Initial Guesses for the Stable and Unstable Manifolds. The eigenvector with eigenvalue greater than 1 is the unstable one; the eigenvector with the eigenvalue less than 1 is the stable one. Let $Y^s(X_0)$ denote the normalized (to 1) stable eigenvector, and $Y^u(X_0)$ denote the normalized unstable eigenvector. We now use these to compute the approximate manifolds as follows: As in equation (4.4.2), let

$$X^s(X_0) = X_0 + \epsilon Y^s(X_0) \tag{7.2.1}$$

be the initial guess for the stable manifold at X_0 along the halo orbit, as illustrated in Figure 4.4.2, and let

$$X^u(X_0) = X_0 + \epsilon Y^u(X_0)$$

be the initial guess for the unstable manifold at X_0. Here ϵ is a small displacement from X_0. The magnitude of ϵ should be small enough to be within the validity of the linear estimate, yet not so small that the time of flight becomes too large due to the asymptotic nature of the stable and unstable manifolds. Gómez, Jorba, Masdemont, and Simó [1991a] suggest values of ϵ corresponding to position displacements of 200 km to 250 km in the Sun-Earth system to get the two branches of the trajectory along the stable and unstable manifolds. By integrating the unstable vector forwards, we generate a trajectory shadowing the unstable manifold. By integrating the stable vector backwards, we generate a trajectory shadowing the stable manifold. For the manifold at $X(t)$, one can simply use the state transition matrix to transport the eigenvectors from X_0 to $X(t)$:

$$Y^s(X(t)) = \Phi(t, t_0; X_0) Y^s(X_0). \tag{7.2.2}$$

Since the state transition matrix does not preserve the norm, the resulting vector must be renormalized.

7.3 Earth-to-Halo Differential Corrector

There are few places along a halo orbit where the corresponding stable manifold passes at the exact altitude of the parking orbit. So, to add flexibility in determining transfer and explore other applications, a differential correction procedure, in combination with a continuation algorithm, is needed for the design process. Various approaches can be used to develop an appropriate corrector. If the manifold and the parking orbit are nowhere

tangent, within the bounds of some specified set of conditions (e.g., time of flight), the problem is redefined as the search for a path between the parking orbit and any point on the manifold. Here, a local approximation of the stable manifold is available near the halo and the corrector solves for the transfer path from the Earth to this position.

Recall from Chapter 6 that the idea of the differential corrector is best described as a process of targeting. Given a reference trajectory $\bar{X}(t)$ going from $\bar{X}_0$ to $\bar{X}_1$ under the natural dynamics $\dot{X} = f(X)$, we want to make slight adjustments, $\delta\bar{X}_0$, of the initial state $\bar{X}_0$ so that the new trajectory will end up at the desired final state X_d. In order to do so, we need to know the sensitivity of the final state $\bar{X}_1$ to small changes in the initial state $\delta\bar{X}_0$. This linear approximation is precisely given by the state transition matrix evaluated along the reference trajectory $\bar{X}(t)$:

$$\delta\bar{X}_1 = \Phi(t_1, t_0)\delta\bar{X}_0. \tag{7.3.1}$$

Recall that Φ is the linearized flow map along the reference trajectory, $\Phi(t_1, t_0) = \frac{\partial\phi(t_1, t_0; \bar{X}_0)}{\partial X}$ and will be denoted below as $\partial\bar{X}_1/\partial\bar{X}_0$ when it is convenient.

For the halo orbit insertion (HOI) problem, a simple differential corrector encapsulated in equation (7.3.1) can be used. This differential corrector used in the computation of Earth-to-halo transfer trajectories was developed in Mains [1993]. For more details, see Barden [1994] and Howell, Mains, and Barden [1994].

The HOI Problem is also called the Launch Problem, or the Earth-to-Halo Transfer Problem, as illustrated in Figure 7.3.1. The question is: How do we find the optimum transfer trajectory from the Earth to the halo orbit? Which is the best HOI point to insert into the halo orbit? That depends on what is meant by the "best", i.e., what are the conditions that we need to optimize. Clearly, one of the major concerns is the magnitude of the HOI ΔV and TTI ΔV (Transfer Trajectory Insertion). Other constraints could be the orbital elements of the Earth parking orbit, the time from TTI to HOI, etc. It all depends on the problem. Hence, there is no unique "best" solution. There are many best solutions, depending on the constraints.

Nevertheless, as ΔV is usually a precious commodity for space missions, minimizing ΔV is important. To that end, the use of the stable manifold is crucial. The question is: How do we get a good initial guess for a near optimal transfer trajectory? Since the stable manifold naturally inserts onto the halo orbit without a maneuver, it provides a very good guess for the HOI transfer trajectory. The problem is, the stable manifold for a particular halo orbit may not get close enough to the Earth to insert onto the manifold. Hence, one condition for a good first guess to the transfer trajectory is to find a halo orbit whose stable manifold comes close to the Earth. Additionally, one needs to find that piece of the stable manifold

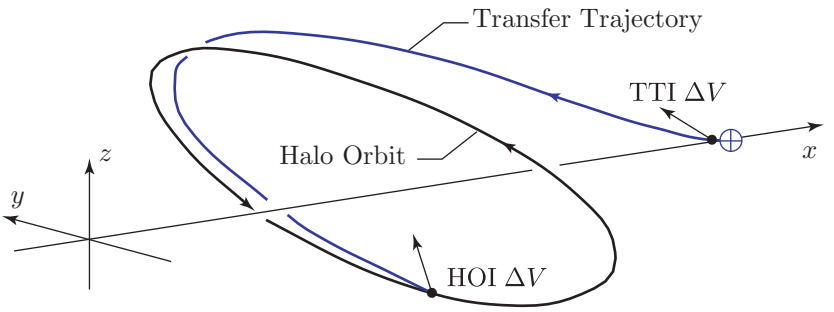


FIGURE 7.3.1. The Launch Problem: finding the best solution from Transfer Trajectory Insertion (TTI) to Halo Orbit Insertion (HOI).

that comes closest to the Earth at the earliest possible instance. We integrate the stable manifold backwards to find these trajectories. Let $X(t)$ be such a trajectory lying on the stable manifold, $X(t_0) = X_0 =$ HOI point, $X(t_1) = X_1 =$ TTI point. But, most likely, X_1 will be unsuitable as a TTI point. Perhaps, the altitude is too high, compared with the 200 km altitude parking orbit used by most launch vehicles. Or the inclination of the start of the transfer trajectory is too far away from the inclination of the parking orbit. We need to differentially correct to find a piece of the suitable stable manifold which approaches our TTI launch constraints.

Select Suitable Variables for the Initial and the Final States. One of the key things in constructing a differential corrector is to select a set of suitable variables which are being physically controlled and which provide stable numerical algorithms. For the Launch Problem, one set of good variables for the initial and the final states are:

$$X_0 = \{x_0, y_0, z_0, \Delta V_{xy}, \beta, \Delta V_z\} = \text{HOI}$$
$$X_1 = \{h_a, \alpha, \lambda, \gamma, V_1, \xi\} = \text{TTI}$$

The initial state X_0 is given in CR3BP coordiantes with origin at the barycenter. Note that the cartesian velocities have been replaced by state elements describing the change in velocity (ΔV) between the transfer trajectory and the nominal orbit at HOI. The vector ΔV is given in ecliptic cylindrical coordinates in the above. The magnitude of the change in velocity in the xy-plane is represented by ΔV_{xy}, and the angle between ΔV_{xy} and the x-axis is denoted by β. Finally, ΔV_z describes the velocity change in the z-direction. These quantities can be expressed in terms of cartesian components as

$$\Delta V_{xy} = [(\dot{x}_0 - \dot{x}_h)^2 + (\dot{y}_0 - \dot{y}_h)^2]^{\frac{1}{2}} \tag{7.3.2}$$
$$\beta = \tan^{-1}\left[\frac{\dot{y}_0 - \dot{y}_h}{\dot{x}_0 - \dot{x}_h}\right] \tag{7.3.3}$$

$$\Delta V_z = \dot{z}_0 - \dot{z}_h, \tag{7.3.4}$$

where the subscript h identifies the velocity components on the nominal path at the HOI, and $\dot{x}_0$, $\dot{y}_0$ and $\dot{z}_0$ are the velocity components on the transfer path at HOI.

The final state X_1 is defined to be near the Earth at the TTI point. The variable h_a is the altitude of the spacecraft above the Earth, and α is the angle between the rotating x-axis and the projection of the position vector r on the xy-plane. λ is the elevation angle, the angle between the position vector r and the xy-plane. γ is the flight path angle (the angle between the velocity and the local horizontal) while V_1 is the magnitude of the velocity. Finally, ξ is the angle between the final plane of motion of the spacecraft and the $xy-$plane (inclination of the launch parking orbit). These quantities can be expressed in terms of cartesian components as follows:

$$h_a = (\bar{x}_1^2 + y_1^2 + z_1^2)^{\frac{1}{2}} - R, \tag{7.3.5}$$

$$\alpha = \tan^{-1}\left(\frac{y_1}{\bar{x}_1}\right), \tag{7.3.6}$$

$$\lambda = \tan^{-1}\left[\frac{z_1}{(\bar{x}_1^2 + y_1^2)^{\frac{1}{2}}}\right], \tag{7.3.7}$$

$$\gamma = \cos^{-1}\left\{\frac{[(y_1\dot{z}_1 + z_1\dot{y}_1)^2 + (z_1\dot{x}_1 - \bar{x}_1\dot{z}_1)^2(\bar{x}_1\dot{y}_1 - y_1\dot{x}_1)^2]^{\frac{1}{2}}}{(h_a + R)(\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2)^{\frac{1}{2}}}\right\}, \tag{7.3.8}$$

$$V_1 = (\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2)^{\frac{1}{2}}, \tag{7.3.9}$$

$$\xi = \cos^{-1}\left[\frac{\bar{x}_1\dot{y}_1 - y_1\dot{x}_1}{(h_a + R)(\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2)^{\frac{1}{2}}}\right], \tag{7.3.10}$$

where R is the radius of the Earth and $\bar{x}_1 = x_1 - 1 + \mu$. The geometry associated with some of these new variables can be seen in Figure 7.3.2.

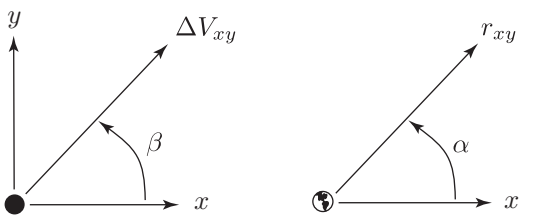


FIGURE 7.3.2. Definition of state elements.

Earth-to-Halo Differential Correction Scheme. Now we need to formulate a strategy: What do we vary? What do we hold fixed? Since we want to get onto the halo orbit, we must fix the position of X_0, i.e., we fix $\{x_0, y_0, z_0\}$. But, we can vary the velocity at HOI. Having determined which initial states can be varied, any constraints on the target states must be defined. Of particular interest are the altitude, flight path angle, and the angle α. The remaining three "final" states are free variables. Using the chain rule, the variations in the TTI states can now be written as a function of the variations in the HOI states,

$$\delta h_a = \frac{\partial h_a}{\partial \Delta V_{xy}}\delta\Delta V_{xy} + \frac{\partial h_a}{\partial \beta}\delta\beta + \frac{\partial h_a}{\partial \Delta V_z}\delta\Delta V_z + \frac{\partial h_a}{\partial \sigma_1}\delta\sigma_1, \quad (7.3.11)$$

$$\delta\gamma = \frac{\partial \gamma}{\partial \Delta V_{xy}}\delta\Delta V_{xy} + \frac{\partial \gamma}{\partial \beta}\delta\beta + \frac{\partial \gamma}{\partial \Delta V_z}\delta\Delta V_z + \frac{\partial \gamma}{\partial \sigma_1}\delta\sigma_1, \quad (7.3.12)$$

$$\delta\alpha = \frac{\partial \alpha}{\partial \Delta V_{xy}}\delta\Delta V_{xy} + \frac{\partial \alpha}{\partial \beta}\delta\beta + \frac{\partial \alpha}{\partial \Delta V_z}\delta\Delta V_z + \frac{\partial \alpha}{\partial \sigma_1}\delta\sigma_1, \quad (7.3.13)$$

where σ_1 is the stopping time at X_1 which is allowed to vary. The above relationships can then be used to produce suitable differential correctors. First consider expressions for the left hand side. Since we want the TTI to occur at perigee, i.e., $\gamma = 0$, for the most effective use of the ΔV, the numerical simulation is terminated when γ achieves its target value (zero). Therefore, $\delta\gamma$ is always assumed to be zero. Now let the desired target end states be h_a^* and α^*. Then the variations are related to the actual states as follows:

$$\delta h_a = h_a^* - h_a, \qquad \delta\alpha = \alpha^* - \alpha.$$

Note that, with the termination conditions on γ fixed, only two states are targeted. To produce a system of two equations and two unknowms, an additional constraint is required on the three velocity states at HOI. It is straightforward to simply fix one of the velocity states at a constant value leaving only two states free to be varied. If ΔV_z is held constant, the result is a relationship between errors in two states at TTI and changes in two HOI velocity states, i.e., a state relation matrix that is expressed in the form

$$\begin{bmatrix} h_a^* - h_a \\ \alpha^* - \alpha \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} \delta\Delta V_{xy} \\ \delta\beta \end{bmatrix},$$

where

$$B_{11} = \frac{\partial h_a}{\partial \Delta V_{xy}} - \frac{\partial h_a/\partial\sigma_1}{\partial\gamma/\partial\sigma_1}\frac{\partial\gamma}{\partial\Delta V_{xy}}$$

$$B_{12} = \frac{\partial h_a}{\partial \beta} - \frac{\partial h_a/\partial\sigma_1}{\partial\gamma/\partial\sigma_1}\frac{\partial\gamma}{\partial\beta}$$

$$B_{21} = \frac{\partial \alpha}{\partial \Delta V_{xy}} - \frac{\partial h_a/\partial\sigma_1}{\partial\gamma/\partial\sigma_1}\frac{\partial\gamma}{\partial\Delta V_{xy}}$$

$$B_{22} = \frac{\partial \alpha}{\partial \beta} - \frac{\partial h_a / \partial \sigma_1}{\partial \gamma / \partial \sigma_1} \frac{\partial \gamma}{\partial \beta}.$$

To obtain B_{ij}, we first solve $\delta\sigma_1$ as a function of $\delta\Delta V_{xy}$ and $\delta\beta$ (with $\delta\gamma = \delta\Delta V_z = 0$) from the equation (7.3.12), then substitute the resulting expression into equations (7.3.11) and (7.3.13) and simplify.

Also, as noted previously, the partial derivatives in the formulas for B_{ij} can be expressed as functions of the state transition matrix along the halo orbit. Let $X_i^R = (x_i, y_i, z_i, \dot{x}_i, \dot{y}_i, \dot{z}_i), i = 0, 1$ be the state vector representing the initial HOI and the final TTI states in the CR3BP coordinates, then the 6×6 state transition matrix Φ can be defined in terms of four 3×3 submatrices

$$\Phi = \begin{bmatrix} \Phi_{rr} & \Phi_{rv} \\ \Phi_{vr} & \Phi_{vv} \end{bmatrix}.$$

Since each of the new variables is a function of the original CR3BP variables (see equations (7.3.2)-(7.3.10)), the partial derivatives of the new variables can be written in terms of the partials of the CR3PB varibles, or Φ_{ij}.

Now take B_{11} as an example. The partial derivatives in its formula can be evaluated from the following expressions:

$$\frac{\partial h_a}{\partial \Delta V_{xy}} = \begin{bmatrix} \frac{\partial h_a}{\partial x_1} & \frac{\partial h_a}{\partial y_1} & \frac{\partial h_a}{\partial z_1} \end{bmatrix} \Phi_{rv} \begin{bmatrix} \frac{\partial \dot{x}_0}{\partial \Delta V_{xy}} \\ \frac{\partial \dot{y}_0}{\partial \Delta V_{xy}} \\ \frac{\partial \dot{z}_0}{\partial \Delta V_{xy}} \end{bmatrix},$$

$$\frac{\partial \gamma}{\partial \Delta V_{xy}} = \begin{bmatrix} \frac{\partial \gamma}{\partial x_1} & \frac{\partial \gamma}{\partial y_1} & \frac{\partial \gamma}{\partial z_1} & \frac{\partial \gamma}{\partial \dot{x}_1} & \frac{\partial \gamma}{\partial \dot{y}_1} & \frac{\partial \gamma}{\partial \dot{z}_1} \end{bmatrix} \begin{bmatrix} \Phi_{rv} \\ \Phi_{vv} \end{bmatrix} \begin{bmatrix} \frac{\partial \dot{x}_0}{\partial \Delta V_{xy}} \\ \frac{\partial \dot{y}_0}{\partial \Delta V_{xy}} \\ \frac{\partial \dot{z}_0}{\partial \Delta V_{xy}} \end{bmatrix},$$

$$\frac{\partial h_a}{\partial \sigma_1} = \begin{bmatrix} \frac{\partial h_a}{\partial x_1} & \frac{\partial h_a}{\partial y_1} & \frac{\partial h_a}{\partial z_1} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{y}_1 \\ \dot{z}_1 \end{bmatrix},$$

$$\frac{\partial \gamma}{\partial \sigma_1} = \begin{bmatrix} \frac{\partial \gamma}{\partial x_1} & \frac{\partial \gamma}{\partial y_1} & \frac{\partial \gamma}{\partial z_1} & \frac{\partial \gamma}{\partial \dot{x}_1} & \frac{\partial \gamma}{\partial \dot{y}_1} & \frac{\partial \gamma}{\partial \dot{z}_1} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{y}_1 \\ \dot{z}_1 \\ \ddot{x}_1 \\ \ddot{y}_1 \\ \ddot{z}_1 \end{bmatrix}.$$

For more details, see Mains [1993].

7.4 Numerical Results

Global Behavior of Stable Manifolds. The first step in the search for Earth-to-halo transfer trajectories is to characterize the behavior of

the globalized stable manifold that orginates from different regions along the halo orbit. Thus, numerical computation of the manifolds associated with various points on the halo orbit is useful. Here, we follow the work of Barden [1994]. With no prior knowledge of which regions are best suited for transfers, some specified number of points are defined along the halo orbit, equally spaced in time. In the case of a northern L_1 halo with A_z =120,000 km, there are 300 points at intervals of 0.59 days. Each point is defined as a fixed point on the halo orbit (points $1, 2, \ldots, k$). The local approximation of the stable manifold is computed for point number 1. The approximation for every other point on the halo can now be calculated by using the state transition matrix between point 1 and the point of interest:

$$Y_i^s = \Phi(t_i, t_1) Y_1^s,$$

where the subscript i indicates the i-th point. For each point, initial state vectors are calculated using equation (7.2.1) and are then use to initiate backward integration to globalize the state manifold. The appropriate pass distance at the Earth is then recorded and labeled case i. Point number 1 is defined as the minimum z point ($y = 0$) and subsequent points are labeled in sequential clockwise order in the xy-projection, the direction of motion along the halo.

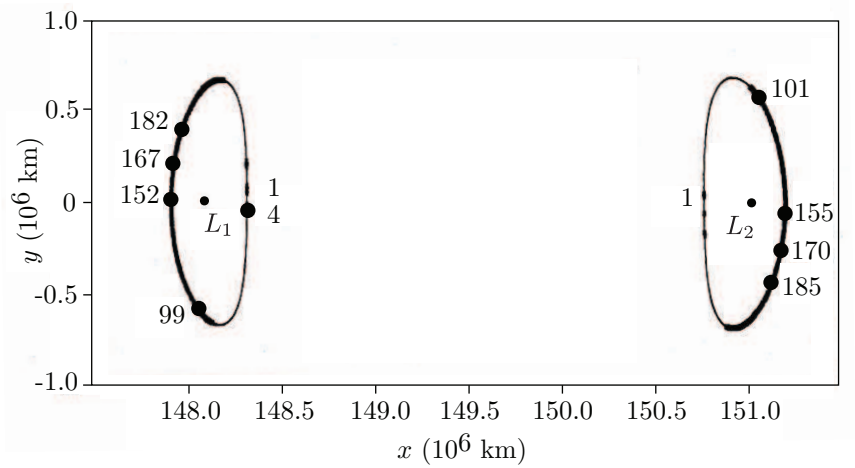


FIGURE 7.4.1. Fixed points that identify stable manifolds passing near Earth for L_1 and L_2 halo orbits (A_z =120,000 km.)

In addition to this halo orbit, a southern halo orbit near L_2 with the same value of out-of-plane amplitude, A_z, and the same number of points is analyzed. For the L_2 orbit, point 1 is defined at the maximum z point and the numbers proceed clockwise along the orbit. For each of these points,

the corresponding stable manifold is propagated backward in time using the method previously described. A critical pass distance from the Earth of 100,000 km is arbitrarily chosen. The results are presented in Figure 7.4.1 for both the L_1 and L_2 halo orbits with A_z =120,000 km. (Note that these transfer are three-dimensional, but the primary characteristics of each transfer can be viewed in the xy-projection. The other two projections are therefore omitted.) The medium-sized dots identify each point associated with a stable manifold that passes the Earth within the critical pass distance. These points would seem to be good candidates to produce the best results in terms of transfer cost. In examing the L_1 output data, the lowest distance on a first pass is approximately 4910 km altitude at point 167, marked with a large dot in Figure 7.4.1. This result is consistent with the example in Gómez, Jorba, Masdemont, and Simó [1991a]. Manifolds in this general region of the nominal halo, that is, between points 152-182, include first Earth passes that are between 4900 and 8700 km altitude. The closest passages along the L_2 orbit are in the region defined by points 155-185, with the lowest, 4960 km at point 170. In general, these regions with the lowest first passage altitudes seem to support comments in Stalos, Folta, Short, Jen, and Seacord [1993] that halo orbit insertion maneuvers may be smaller near the "halo apogee."

Ideally, the "best" result would be to find a stable manifold that passes at an altitude exactly equal to that of the desired parking orbit. For the halo orbit seen in Figure 7.4.1, no such manifolds are found, at least on the first pass by the Earth. However, two points are identified with corresponding manifolds that possess the desired altitude with respect to the Earth on a pass beyond the first. After multiple passes, the stable manifold associated with point 99 on the L_1 orbit, passes the Earth at an altitude of 185 km. The resulting xy-projection of the path can be seen in Figure 7.4.2. The result for point 99 is significant because it is a possible transfer with no HOI cost: a free transfer. The obvious trade off is the time of flight (TOF); for this transfer, TOF is approximately 336 days (as compared to 114 days for ISEE-3). The other stable manifold that results in a path reaching an altitude of 185 km, and also occurs after multiple passes, is associated with point 4. The TOF for this transfer is over 778 days which makes practical use doubtful. However, this particular transfer is extremely useful in suggesting new solutions such as a halo-to-halo transfer between different libration points and it is analyzed in greater detail later in Barden [1994].

The same analysis can be performed for slightly larger amplitude halo orbits with the same corresponding region appearing to be favorable, e.g., "halo apogee". In fact, for slightly larger amplitude halo orbits, there are pieces of the manifold which reach the desired low altitudes (185 km) on their first pass, yielding free transfers with TOFs on the order of only 100 days.

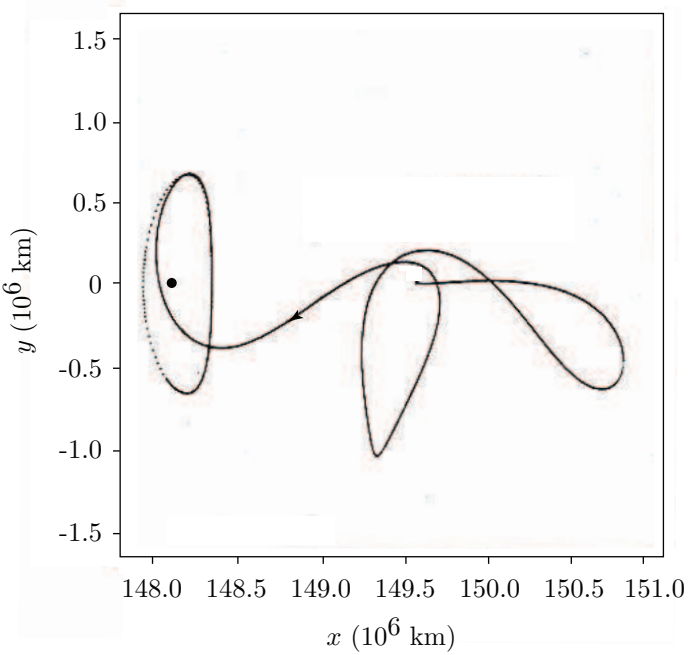


FIGURE 7.4.2. Transfer near fixed point 99; no halo insertion cost.

A Differential Correction Process to Generate Transfer Trajectories. For various resons, it may be necessary to target a specific point on or near a particular halo orbit. Of course, the stable manifold associated with such a fixed point generally does not pass the Earth at a reasonable altitude for a parking orbit. Thus differential correction is used to produce a transfer. This final step begins with the globalized stable manifold and corrects on the velocity at X_0 based on differences between the computed and desired "final" states at TTI. The need for a continuation algorithm becomes clear when attempting to produce an ISEE-3 type transfer to the minimum z point on the halo. The stable manifold first passes the Earth at an altitude of approximately 550,000 km, far beyond the linear range if it is desired to target an altitude of 185 km. Instead of targeting an altitude of 185 km initially, however, another target altitude ($h^{(1)}$) is chosen that is slightly less than 550,000 km. Upon converging to a solution that reaches $h^{(1)}$, the resulting trajectory is then used as an initial guess for a new, lower altitude ($h^{(2)}$). This is repeated until a transfer is determined that originates near an altitude of 185 km; the final transfer can thus be determined. The continuation process can be visualized in Figure 7.4.3 where intermediate transfers are bounded by the initial path generated from the stable manifold and the final transfer trajectory with an HOI cost of 32.3 m/s.

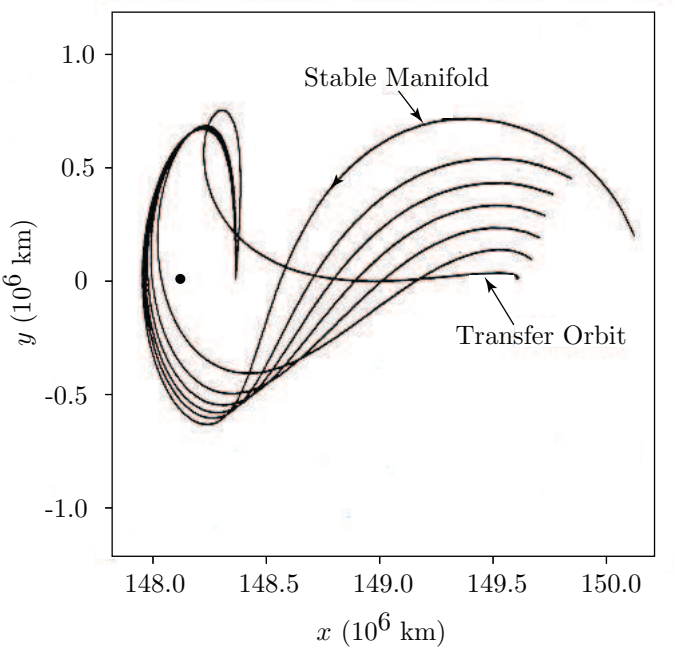


FIGURE 7.4.3. Example of continuation.

As was mentioned, no point along the L_1 halo of amplitude A_z =120,000 km corresponds to a 185 km altitude Earth passage distance at its first perigee. Yet, a direct transfer from Earth to the halo orbit may be desirable. (A *direct transfer* is defined such that backwards integration along the transfer path from the halo reaches the Earth parking orbit, or TTI, on the first pass.) To construct a direct transfer that may be less costly than the result in Figure 7.4.3 an alternative is to choose any point with a manifold that passes closer to the Earth. The stable manifold associated with the previously mentioned point 167 passes closest at 4910 km. Using this stable manifold as a first guess for the differenitial corrector, a transfer trajectory can be determined. See Figure 7.4.4. It is noted that this result requires 20.3 m/s for insertion. Further improvement in the cost may be possible by adjusting some the parameters used in the definition of the target point and the computation of the manifold.

7.5 Return Trajectory Differential Correction

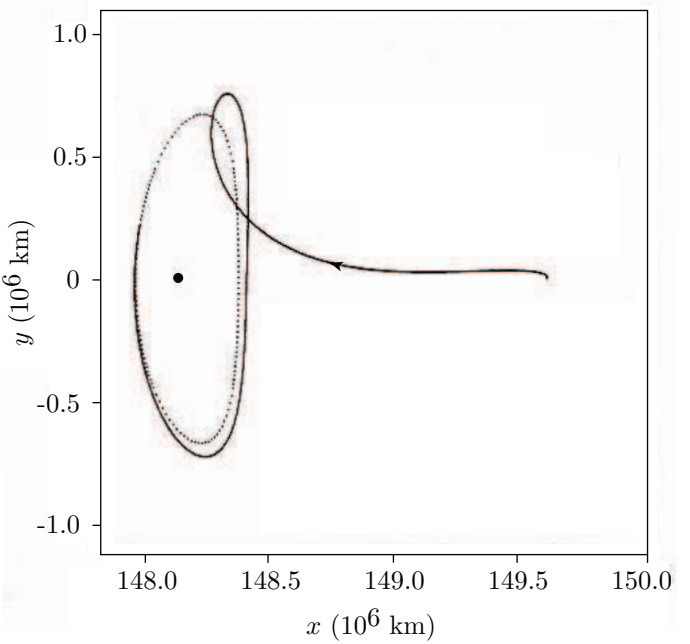


FIGURE 7.4.4. Transfer near fixed point 167; halo insertion cost = 20.3 m/s.

Frequently, a simple differential corrector such as the HOI corrector above may not work. The trajectory is too "stiff". The initial condtions are too sensitive. Or there are too many constraints at the end point than there are parameters at the starting point. In these cases, we simply break the trajectory into segments by adding additional ΔV's. The places where two different trajectory segments are patched together are called *patch points*. In this instance, we are using multiple shooting, instead of simple shooting.

But straight multiple shooting is frequently very sensitive and convergence becomes a problem unless an extremely good initial guess is provided, which is frequently hard to come by. A variant which we call ***two-level differential correction*** proceeds as follows. We start with multiple segments of trajectories which may be discontinuous in position and velocity.[1]

Level I Differential Corrector. We vary the velocities at the patch points to correct the discontinuity of position using the state transition matrix. For this corrector, we work on one segment at a time, even though there are multiple segments. Usually, only one iteration is performed. Hence the position discontinuity is reduced, corrected to first order. Constraints

[1]We thank Roby Wilson for sharing his unpublished notes on the differential corrector for the Genesis trajectory, on which this section is based.

are not applied at this level.

Let us describe the procedure in more detail. The Level I differential corrector is based on only two patch points: an initial state and a final target state (see Figure 7.5.1). The goal is to determine changes in the initial velocity such that, after appropriate propagation, the final position and time is achieved. This is essentially the solution of a two-point boundary value problem.

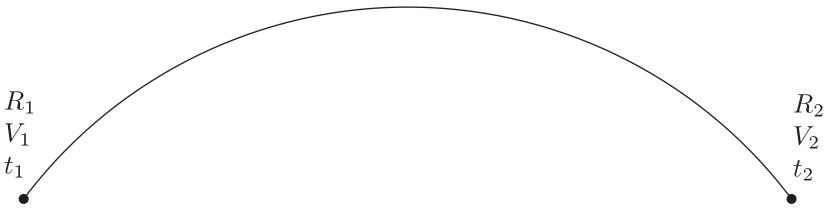


FIGURE 7.5.1. Level I Differential Corrector: two patch points.

The following linear system represents the relationship between changes in the state at one patch and changes in the other patch point,

$$\begin{bmatrix} \delta R_2 - V_2 \delta t_2 \\ \delta V_2 - a_2 \delta t_2 \end{bmatrix} = \begin{bmatrix} A_{21} & B_{21} \\ C_{21} & D_{21} \end{bmatrix} \begin{bmatrix} \delta R_1 - V_1 \delta t_1 \\ \delta V_1 - a_1 \delta t_1 \end{bmatrix}, \tag{7.5.1}$$

where the vector (t_1, R_1, V_1, a_1) represents the state at the initial time, and (t_2, R_2, V_2, a_2) represents the state at the final time. Here, $R_i = (x_i, y_i, z_i), V = (\dot{x}_i, \dot{y}_i, \dot{z}_i), a_i = (\ddot{x}_i, \ddot{y}_i, \ddot{z}_i), i = 1, 2$. The subscripts on the submatrices denote the direction of the state transition matrix. For example, A_{12}, represents changes in the position of state 1 due to changes in the postion at state 2, while C_{21} represents changes in the velocity of state 2 due to changes in the position of state 1.

The formulation for the Level I differential corrector can be described by the following five terms,

$$\text{dimensions} : 14, \tag{7.5.2}$$

$$\text{fixed constraints} : \delta R_1 = 0;\ \delta t_1 = \delta t_2 = 0, \tag{7.5.3}$$

$$\text{targets} : \delta R_2 = -\Delta R_2 = R_{2d} - R_2, \tag{7.5.4}$$

$$\text{controls} : \delta V_1, \tag{7.5.5}$$

$$\text{free} : \delta V_2. \tag{7.5.6}$$

The dimension is the total number of parameters available in the problem. Since there are only two states (epoch, position, velocity) there are $2 \times 7 = 14$ total dimensions. Notice that acceleration is not included because of the constraint $\delta t_1 = \delta t_2 = 0$. The fixed constraints are the parameters in the problem that are not allowd to vary. In this case the initial position is fixed so the variation in R_1 is zero. The initial and final times are also fixed so no

variation is allowed in these quantities. The targets describes the target or goal of the differential correction process. In this case, we are attempting to achieve a prescribed final position, therefore the variation in final position (δR_2) is set to be the error between the desired position (R_{2d}) and the actual position (R_2). The controls are the parameters that are available to be varied to meet the targets. In this case, the initial velocity (V_1) is used to target the final position. Finally, the free parameters are those that do not enter into the differential correction process. In this case, the final velocity (V_2) is free. Note that the degrees of freedom from each of the fixed constraints, targets, controls, and free parameter must equal the number of dimensions of the problem.

From this description of the process, it is necessary to determine a relationship between the targets and the controls, while meeting the fixed constraints. Applying the constraints in equation (7.5.3) to the variations in equation (7.5.1) results in the following relationship.

$$\begin{bmatrix} \delta R_2 \\ \delta V_2 \end{bmatrix} = \begin{bmatrix} A_{21} & B_{21} \\ C_{21} & D_{21} \end{bmatrix} \begin{bmatrix} 0 \\ \delta V_1 \end{bmatrix}, \tag{7.5.7}$$

from which, the targets (R_2) and controls (V_1) can be related by

$$\delta R_2 = B_{21} \delta V_1. \tag{7.5.8}$$

This variational equation can be used to determine the changes in initial velocity required to reduce the changes in final postion, that is,

$$\delta V_1 = B_{21}^{-1} \delta R_2, \tag{7.5.9}$$

or

$$\Delta V_1 = B_{21}^{-1} \Delta R_2, \tag{7.5.10}$$

where Δ now denotes actual changes in the state of the patch points. If this change in initial velocity (ΔV_1) is applied to current initial state and propagated forward to the final state, the error in final position (ΔR_2) should be smaller than the previous iteration. Note that δR_2 is chosen as $-\Delta R_2$ in order to reduce the velocity error ΔV_1. Since the relationships developed are linear, while the propagation is decidedly non-linear, this is an iterative process which usually converges to within a desired tolerance within a few iterations.

Note that a similar relationship can be derived by applying the constraints in equation (7.5.3) to the inverse of equation (7.5.1). This results in the following relationship

$$\begin{bmatrix} 0 \\ \delta V_1 \end{bmatrix} = \begin{bmatrix} A_{12} & B_{12} \\ C_{12} & D_{12} \end{bmatrix} \begin{bmatrix} \delta R_2 \\ \delta V_2 \end{bmatrix}. \tag{7.5.11}$$

From this relationship, two variational equations can be determined. By eliminating δV_2 and solving for δR_2 in terms of δV_1, the relationship

$$\delta R_2 = [C_{12} - D_{12} B_{12}^{-1} A_{12}]^{-1} \delta V_1, \tag{7.5.12}$$

between the controls and targets is derived. This expression is equivalent to the one in equation (7.5.9). In general, the most compact form will be used with the understanding that other equivalent forms could be used interchangeably. We now proceed to the Level II Corrector.

Level II Differential Corrector. The Level II differential corrector is a procedure based on three or more patch points. The goal of the Level II corrector is to simultaneously determine some set of changes for all of the patch points to meet some given set of targets. The first patch point is called the initial state, the last patch point is called the final state, and all the rest are termed interior patch points. The set of patch points define trajectory segments (or legs) between consecutive patch points. Hence, if there are n patch points, there are $n-1$ trajectory segments connecting the patch points.

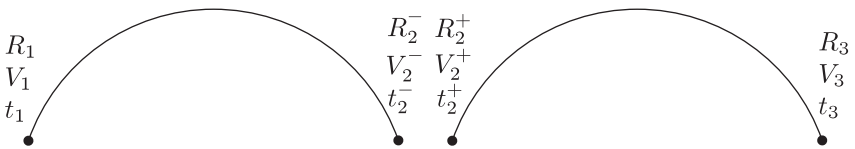


FIGURE 7.5.2. Level II Differential Corrector: 3 patch points.

An example problem with three patch points is represented in Figure 7.5.2 where subscript 1 denotes the initial state, 2 is the interior state, and 3 is the final state. For interior states, the state at the end of the propagation from the previous patch point is given the superscript $(-)$ and is called the incoming state. Conversely, the outgoing state at the beginning of the next trajectory leg is denoted with a $(+)$ supersript.

For this particular configuration, the following linear relationships between state variations are valid,

$$\begin{bmatrix} \delta R_2^- - V_2^- \delta t_2^- \\ \delta V_2^- - a_2^- \delta t_2^- \end{bmatrix} = \begin{bmatrix} A_{21} & B_{21} \\ C_{21} & D_{21} \end{bmatrix} \begin{bmatrix} \delta R_1 - V_1 \delta t_1 \\ \delta V_1 - a_1 \delta t_1 \end{bmatrix}, \quad (7.5.13)$$

$$\begin{bmatrix} \delta R_2^+ - V_2^+ \delta t_2^+ \\ \delta V_2^+ - a_2^+ \delta t_2^+ \end{bmatrix} = \begin{bmatrix} A_{23} & B_{23} \\ C_{23} & D_{23} \end{bmatrix} \begin{bmatrix} \delta R_3 - V_3 \delta t_3 \\ \delta V_3 - a_3 \delta t_3 \end{bmatrix}. \quad (7.5.14)$$

Note that for this problem, each trajectory segment is treated as independent with respect to the rest of the segments in the solution. This means, for example, the variations in the initial state (at point 1) do not directly affect changes on the second segment from 2^+ to 3. This is a critical assumption that will be explained momentarily.

Using the same terminology as before, the formulation for this particular Level II differential corrector is

dimensions : 28,

$$\begin{aligned}
\text{fixed constraints}: & R_2^- = R_2^+ = R_2;\ t_2^- = t_2^+ = t_2,\\
\text{targets}: & \delta V_2^-, \delta_2 V^+,\\
\text{(actually)} \quad & \delta\Delta V_2 = \delta V_2^+ - \delta V_2^- = \Delta V_d - (V_2^+ - V_2^-),\\
\text{controls}: & \delta R_1, \delta t_1, \delta R_2, \delta t_2, \delta R_3, \delta t_3,\\
\text{free}: & \delta V_1, \delta V_3.
\end{aligned}$$

As before, dimensions represents the number of parameters available in the problem. In general, for the Level II corrector the number of dimensions is equal to $2(n-1)\times 7$ representing two states for each trajectory segment (counting incoming and outgoing states as distinct). The fixed constraints represent the continuity constraints at all of the interior patch points. This ensures that the trajectory is continous in position and time (via application of the Level I differential corrector). In general, there are $(n-1)\times 4$ fixed constraints. It is these constraints on position and time that allows trajectory segments to be treated as independent, thus isolating the variational effects of each state's changes to only one segment. If these fixed constraints are changed, the Level II differential corrector must be re-derived. The targets for the Level II corrector are the velocities at the interior patch points. More specifically, the velocity discontinuity ($\Delta V_2 = V_2^+ - V_2^-$) between the incoming velocity and the outgoing velocity is targeted to a specified value. If a fully continuous trajectory is desired, then the targeted value (ΔV_d) is zero. If a deterministic maneuver is allowed at this patch point, then the desired value can be some non-zero vector. There are $(n-2)\times 3$ such targets, 3 for each interior patch point. Additionally, it is possible to specify arbitrary constraints at any of the patch points. These constraints become additional targets. The discussion of these constraints is detailed in the next section. For the Level II corrector, the controls are the positions and times of all the patch points in the problem including the initial and final states. So, in general, there are $n\times 4$ control parameters. Note that unless there are too many additional constraints added to the targets, this system will be naturally underdetermined. Finally, the free parameters in this formulation are the initial and final velocities (V_1 and V_3 respectively).

To solve the 3 patch point problem, the following variations are necessary relating the targets (ΔV_2) to the controls ($\delta R_1, \delta t_1, \delta R_2, \delta t_2, \delta R_3, \delta t_3$), i.e.,

$$\frac{d\Delta V_2}{dR_i}; \qquad \frac{d\Delta V_2}{dt_i}; \qquad i = 1,2,3. \tag{7.5.15}$$

By using similar procedures as in Level I, we obtain

$$\begin{aligned}
\frac{d\Delta V_2}{dR_1} &= -\frac{\partial V_2^-}{\partial R_1} = -B_{12}^{-1},\\
\frac{d\Delta V_2}{dt_1} &= -\frac{\partial V_2^-}{\partial t_1} = -B_{12}^{-1} V_1,
\end{aligned}$$

$$\frac{d\Delta V_2}{dR_2} = \frac{\partial V_2^+}{\partial R_2} - \frac{\partial V_2^-}{\partial R_2} = -B_{32}^{-1}A_{31} + B_{12}^{-1}A_{12},$$
$$\frac{d\Delta V_2}{dt_2} = \frac{\partial V_2^+}{\partial t_2} - \frac{\partial V_2^-}{\partial t_2} = a_2^+ - a_2^- + B_{32}^{-1}A_{32}V_2^+ - B_{12}^{-1}A_{12}V_2^-,$$
$$\frac{d\Delta V_2}{dR_3} = \frac{\partial V_2^+}{\partial R_3} = B_{32}^{-1},$$
$$\frac{d\Delta V_2}{dt_3} = \frac{\partial V_2^-}{\partial t_3} = -B_{32}^{-1}V_3.$$

So the complete relationship describing the Level II differential corrector for this setup is

$$\delta\Delta V_2 = M \begin{bmatrix} \delta R_1 \\ \delta t_1 \\ \delta R_2 \\ \delta t_2 \\ \delta R_3 \\ \delta t_3 \end{bmatrix}, \tag{7.5.16}$$

where

$$M = \begin{bmatrix} \frac{d\Delta V_2}{dR_1} & \frac{d\Delta V_2}{dt_1} & \frac{d\Delta V_2}{dR_2} & \frac{d\Delta V_2}{dt_2} & \frac{d\Delta V_2}{dR_3} & \frac{d\Delta V_2}{dt_3} \end{bmatrix}, \tag{7.5.17}$$

is the matrix containing all of the partial derivatives and is given by the name *state relationship matrix* or SRM. This linear system is underdetermined, that is, there are more controls than targets. Hence, there are infinitely many solutions to this linear system. Choosing the smallest Euclidean norm produces a solution to this system given by

$$\begin{bmatrix} \delta R_1 \\ \delta t_1 \\ \delta R_2 \\ \delta t_2 \\ \delta R_3 \\ \delta t_3 \end{bmatrix} = M^T(MM^T)^{-1}\delta\Delta V_2 \tag{7.5.18}$$

Note that $M^T(MM^T)^{-1}$ is commonly known as a *pseudo-inverse*. By selecting $\delta\Delta V_2 = \Delta V_d - (V_2^+ - V_2^-)$, a set of changes can be determined to the patch point positions and times. These changes are added to the patch point states and the trajectory is re-coverged using a Level I differential correction process. The resulting solution should have a lower target cost (ΔV_2 in this case) than the previous iteration.

Level II with Constraints Often times it is necessary to specify additional constraints on a particular trajectory solution. The Level II differential corrector described in the previous section can be modified to allow

constraints at any of the patch points that describe the solution. This is possible so long as the constraint is of the form

$$\alpha_{ij} = \alpha_{ij}(R_i, V_i, t_i),$$

that is, the constraint α_{ij} is a function of position, velocity, and time of the patch point. The first subscript index i on the constraint denotes the patch point that the constraint is associated with, while the second index j denotes the constraint number at that patch point. This allows for multiple constraints at multiple patch points.

Functionally, the constraints are additional targets in the terminology used to describe the differential corrector. Thus, it is necessary to determine the variations of the desired constraints relative to the controls present in the solution procedure. These variations are placed in the SRM relating targets to the controls. This new relationship can be written as

$$\begin{bmatrix} \delta\Delta V_i \\ \delta\alpha_{ij} \end{bmatrix} = \begin{bmatrix} \frac{d\Delta V_i}{dR_k} & \frac{d\Delta V_i}{dt_k} \\ \frac{d\alpha}{dR_k} & \frac{d\alpha}{dt_k} \end{bmatrix} \begin{bmatrix} \delta\Delta R_k \\ \delta t_k \end{bmatrix} = \bar{M} \begin{bmatrix} \delta\Delta R_k \\ \delta t_k \end{bmatrix}$$

where the matrix $\bar{M}$ is the *augumented SRM* that contains the velocity and constraint variables.

The derivation of the constraint variations with respect to the patch points is consistent with the derivation from the previous section relating the velocity discontinuities to the patch points. Thus, a given constraint has variations from at most the patch points on the trajectory segments immediately before and after, as well as the patch point itself where the constraint is defined. Hence, a given constraint α_{ij} requires the following total derivatives

$$\frac{d\alpha_{ij}}{dR_k}, \qquad \frac{d\alpha_{ij}}{dt_k} \qquad \text{for} \quad k = (i-1), i, (i+1) \tag{7.5.19}$$

at a particular patch point i.

As an example, consider the constraint on radial distance from a reference. In this case, the constraint is given by

$$\alpha_{ij} = |R_i| - R_d,$$

where R_d is the desired radial distance. Since the constraint is only specified in terms of the independent variable R_i, the only non-zero derivaive is

$$\frac{d\alpha}{dR_i} = \frac{\partial\alpha}{\partial R_i} = \frac{R_i^T}{|R_i|}.$$

For constraints that depend on velocity, the situation becomes more complex. In the formulation of the specific Level II differential corrector from the previous section, velocities depend on the positions and times of the

patch points (via the Level I differential correction process). Thus the total derivatives in equation (7.5.19) can be expressed as

$$\frac{d\alpha_{ij}}{dR_k} = \frac{\partial \alpha_{ij}}{\partial R_k} + \frac{\partial \alpha_{ij}}{\partial V_k^-}\frac{\partial V_k^-}{\partial R_k} + \frac{\partial \alpha_{ij}}{\partial V_k^+}\frac{\partial V_k^+}{\partial R_k} \tag{7.5.20}$$

$$\frac{d\alpha_{ij}}{dt_k} = \frac{\partial \alpha_{ij}}{\partial t_k} + \frac{\partial \alpha_{ij}}{\partial V_k^-}\frac{\partial V_k^-}{\partial t_k} + \frac{\partial \alpha_{ij}}{\partial V_k^+}\frac{\partial V_k^+}{\partial t_k}. \tag{7.5.21}$$

Take as an example the simple end-state (R_N, t_N, V_N) constraint that are formally stated as

$$\begin{aligned} C_{N1} &= R_N - R_d \\ C_{N2} &= t_N - t_d \\ C_{N3} &= V_N - V_d \end{aligned}$$

we have the following related variations

$$\begin{aligned} \frac{dC_{N1}}{dR_N} &= I_3 \\ \frac{dC_{N2}}{dt_N} &= 1 \\ \frac{dC_{N3}}{dR_N} &= B_{N-1,N}^{-1} A_{N-1,N} \\ \frac{dC_{N3}}{dt_N} &= -a_N - B_{N-1,N}^{-1} A_{N-1,N} V_N \\ \frac{dC_{N3}}{dR_{N-1}} &= -B_{N-1,N}^{-1} \\ \frac{dC_{N3}}{dt_{N-1}} &= B_{N-1,N}^{-1} V_N \end{aligned}$$

where I_3 is the 3×3 identity matrix. By appropriate choice of the independent parameters, any constraint functions active in the solution can be driven to zero, thus enforcing the constraints. For more details, see Wilson and Howell [1998].

8

Transfer Correction Maneuvers and Station-Keeping Strategies

8.1 Introduction

When initially computing an impulsive burn for spacecraft trajectory, one makes the idealistic assumption that the rocket on the vehicle can deliver the exact amount of thrust necessary to boost the velocity of the spacecraft the amount ΔV required by orbital dynamics. In practice, deviations from the required ΔV are introduced by inaccuracies in the rocket engine which must be corrected by subsequent impulsive maneuvers. Furthermore, correction may be necessary to compensate for effects, gravitational or otherwise, which are not taken into account by dynamical models.

We can separate correction maneuvers into two categories. The first, which are called ***transfer correction maneuvers***, occur during the transfer phase of the space mission, e.g., transfer from the Earth to a Sun-Earth libration point orbit. The second category of correction maneuvers are those which keep a spacecraft on a desired periodic orbit or within a bounded region of the phase space, e.g., a Lissajous orbit around the Sun-Earth libration point. These are called ***station-keeping maneuvers***. We will cover both kinds of correction maneuvers within the context of libration point missions in this chapter. Due to the sensitivity of trajectories near the libration points, correction maneuvers are necessary for the success of any libration point mission.

Transfer Correction Maneuvers Using Optimal Control. The first part of this chapter (§8.2-8.5) addresses the computation of the required trajectory correction maneuvers for a halo orbit space mission to compen-

sate for the launch velocity errors introduced by inaccuracies of the launch vehicle. By combining dynamical systems theory with optimal control techniques, we are able to provide a compelling portrait of the complex landscape of the trajectory design space. This approach enables automation of the analysis to perform parametric studies that simply were not available to mission designers a few years ago, such as how the magnitude of the errors and the timing of the first transfer correction maneuver affects the correction ΔV. The impetus for combining dynamical systems theory and optimal control in this problem arose in the context of design issues for the Genesis Discovery Mission (see supplement in §1.2).

Station-Keeping Strategies. Since periodic and quasi-periodic libration point trajectories are unstable, spacecraft moving on such orbits need control to remain close to their nominal orbit. The second part of this chapter (§8.6-8.7) explores two particular station-keeping strategies for a halo orbit mission using the Target Point strategy and the Floquet Mode approach. For details, see Gómez, Howell, Masdemont, and Simó [1998a].

Halo and Lissajous orbits are solutions that represent bounded motion in the vicinity of a collinear libration point. The problem of controlling a spacecraft moving near an inherently unstable libration point orbit is important. In the late 1960s Farquhar suggested several station-keeping strategies for nearly-periodic solutions near the collinear points (Farquhar [1966, 1968, 1969]). Later, a station-keeping method for spacecraft moving on halo orbits in the vicinity of the Earth-Moon translunar libration point was developed by Breakwell, Kamel, and Ratner [1974]. These studies assumed that the control could be modelled as continuous. In contrast, specific mission requirements led to an impulsive burn station-keeping strategy for the first actual libration point mission. Launched in 1978, the ISEE-3 spacecraft, mentioned in §6.2, remained in a near-halo orbit associated with the interior libration point, L_1, of the Sun-Earth/Moon barycenter system for approximately three and half years. Impulsive maneuvers at discrete time intervals (up to 90 days) were successfully implemented as a means of trajectory control. The ΔV required was minimal. Since that time, more detailed investigations have resulted in various station-keeping strategies, including the two discussed here as the Target Point and Floquet Mode approaches.

The Target Point method (as presented by Howell and Pernicka [1988], Howell and Gordon [1994], and Keeter [1994]) computes corrective maneuvers by minimizing a weighted cost function. The cost function is defined in terms of a corrective maneuver as well as position and velocity deviations from a nominal orbit at a number of specified future times t_i. The nominal state vectors at each time t_i are called "target points." The target points are selected along the trajectory at discrete time intervals that are downstream of the maneuver. In contrast, the Floquet Mode approach, as developed by Simó, Gómez, Llibre, and Martinez [1986], incorporates

invariant manifold theory and Floquet modes to compute the maneuvers. Floquet modes associated with the monodromy matrix are used to determine the unstable component corresponding to the local vector error. The maneuvers are then computed to eliminate the dominant unstable component of the error. Both approaches can maintain control of the vehicle over the duration of the simulation. Although the Target Point approach is more straightforward, a clear advantage in favor of the Floquet Mode approach is the qualitative information and analysis that becomes available. This may be more critical in the Earth-Moon system where the timescale for instability is short. In any case, some combination of these ideas is likely to improve the results. The simplicity and robustness of the Target Point method can complement the qualitative advantages of the Floquet Mode approach.

8.2 Halo Orbit Mission Correction Maneuvers

We will continue using Genesis as an example for the generic prodecure outlined here. In this chapter, we will assume that the halo orbit, $H(t)$, and the stable manifold $M(t)$ are fixed and provided. Hence we will not dwell further on the theory of their computation which is well covered in earlier chapters and other references (see Howell, Barden, and Lo [1997]). Instead, let us turn our attention to the transfer correction maneuver (TCM) problem.

Our main goal in this chapter is to show the feasability of merging optimal control software with dynamical systems methods to compensate for launch vehicle errors, to which libration point missions are particularly sensitive. For this task, we have used a particular piece of software, COOPT, as our demonstration tool. There are many other good software packages which could have been used, such as MISER, used in Liu, Teo, Jennings, and Wang [1998] and Rehbock, Teo, Jennings, and Lee [1999], NTG (Nonlinear Trajectory Generation) of Milam [2003], and the iterative dynamic programming approach of Luus [2000]. However, our goal is not to compare software packages. Rather, it is to combine dynamical systems theory with optimal control techniques to solve a sophisticated problem for an actual space mission.

The Transfer Correction Maneuver. Genesis was launched in 2001 from a Delta 7326 launch vehicle using a Thiakol Star37 motor as the final upper stage. The most important error which was expected by the inaccuracies of the launch vehicle was the velocity magnitude error. In this case, the expected error is 7 m/s (1 sigma value) relative to a velocity boost of approximately 3200 m/s from a 200 km circular altitude Earth orbit.

Although a 7 m/s error for a 3200 m/s maneuver may seem small, it actually is considered quite large. Unfortunately, one of the characteristics of halo orbit missions is that, unlike interplanetary mission launches, they are extremely sensitive to launch errors. Typical interplanetary launches can correct launch vehicle errors 7 to 14 days after the launch. In contrast, halo orbit missions must generally correct the launch error within the first 7 days after launch or the ΔV required to correct will increase beyond the spacecraft's propulsion capability. The most important transfer correction maneuver is called TCM1, being the first TCM of the mission. Two 'clean up' maneuvers, TCM2 and TCM3, generally follow TCM1 after a week or more, depending on the situation.

In the early stages of the launch, the spacecraft is in an elliptical orbit around the Earth. Thus, using conic models in this phase is appropriate. From the energy equation for a spacecraft in a conic orbit about the Earth,

$$E = \frac{V^2}{2} - \frac{GM}{R}, \tag{8.2.1}$$

where E is Keplerian energy, V is velocity, GM is the gravitational mass of the Earth, and R is the distance of the spacecraft from the Earth's center, it follows that

$$\delta V = \frac{\delta E}{V}, \tag{8.2.2}$$

where δV and δE denote the variations in velocity and energy, respectively. The launch velocity error imparts an energy error δE to the spacecraft transfer orbit. In particular, for a highly elliptical orbit such as the one for Genesis soon after launch (see Figure 1.2.6) , V decreases sharply as a function of time since launch. Hence, the magnitude of the correction maneuver ΔV required to cancel the resulting δV grows sharply with time since launch. For a large launch vehicle error, which is possible in the case of Genesis, the correction maneuver TCM1 can quickly grow beyond the capability of the spacecraft's propulsion system.

The computation of TCMs is performed on the ground, relying upon accurate knowledge of the spacecraft's position and velocity. The time necessary for initial spacecraft checkout procedures, which frequently require several days after launch, compel us to investigate the effect of delays in the timing of TCM1. To thoroughly check out the spacecraft's position, velocity, and condition, it is desirable to delay TCM1 by as long as possible, even at the expense of an increased ΔV for TCM1. Consequently, the Genesis navigation team prefered that TCM1 be performed at 2 to 7 days after launch, or later if at all possible.

The design of the Genesis TCM1 retargeted the state after launch back to the nominal Halo Orbit Insertion (HOI) state (see Lo, Williams, Bollman, Han, Hahn, Bell, Hirst, Corwin, Hong, Howell, Barden, and Wilson [2001]). This approach is based on linear analysis and is adequate only if TCM1 were to be performed within 24 hours after launch. But as emphasized,

this was not the case for Genesis, which could permit a TCM1 only after 24 hours past launch, and probably more like 2 to 7 days past launch. See also Wilson, Howell, and Lo [1999] for another approach to targeting that may be applicable for a mission like Genesis.

The requirement of a relatively long time delay between launch (when error is imparted) and TCM1 (when error is corrected) suggests that one use a nonlinear approach, combining dynamical systems theory with optimal control techniques. We explore two similar but slightly different approaches and are able to obtain in both cases an optimal maneuver strategy that fits within the Genesis ΔV budget of 150 m/s for the transfer portion of the trajectory. These two approaches are:

1. Halo Orbit Insertion (HOI) technique: use optimal control techniques to retarget the halo orbit with the original nominal trajectory as the initial guess.

2. Stable Manifold Orbit Insertion (MOI) technique: target the stable manifold.

Both methods are shown to yield good results. We cover these methods in §8.3-8.5, which come from Serban, Koon, Lo, Marsden, Petzold, Ross, and Wilson [2002].

8.3 Optimal Control for Transfer Correction Maneuvers

We now introduce the general problem of optimal control for the spacecraft trajectory planning problem. We start by recasting the TCM problem as a spacecraft trajectory planning problem. Mathematically they are exactly the same. We discuss the spacecraft trajectory planning problem as an optimization problem and highlight the formulation characteristics and particular solution requirements. Then the loss in fuel efficiency caused by possible perturbation in the launch velocity and by different delays in TCM1 is exactly the sensitivity analysis of the optimal solution. COOPT, the software we use, is described later in this section. It is an excellent tool in solving this type of problem, both in providing a solution for the trajectory planning problem with optimal control, and in studying the sensitivity of different parameters (Serban and Petzold [2001]).

We emphasize that the objective in this part of the chapter is not to design the original nominal transfer trajectory, but rather to investigate recovery issues related to possible launch velocity errors which cause the spacecraft to deviate from the nominal trajectory. We therefore assume that a nominal transfer trajectory (corresponding to zero errors in launch

velocity) is available. For the nominal trajectory in our numerical experiments in this chapter, we do not use the actual Genesis mission transfer trajectory, but rather an approximation obtained with a more restricted model, which has a single maneuver of about 13 m/s at the point of insertion onto the halo orbit. It has been shown by Howell, Barden, and Lo [1997] that the general qualitative characteristics found in the restricted models translate well when extended into more accurate models; we expect the same correlation with this work as well.

Recasting TCM as a Trajectory Planning or Rendezvous Problem. Although different from a dynamical systems perspective, the HOI and MOI problems are very similar once cast as optimization problems. In the HOI problem, a final maneuver (jump in velocity) is allowed at $T_{\text{HOI}} = t_{\max}$, while in the MOI problem, the final maneuver takes place on the stable manifold at $T_{\text{MOI}} < t_{\max}$ and no maneuver is allowed at $T_{\text{HOI}} = t_{\max}$. A halo orbit insertion trajectory design problem can be simply posed as the following.

> **Statement of the problem**: *Find the maneuver times and sizes to minimize fuel consumption (ΔV) for a trajectory starting near Earth and ending on the specified halo orbit around the libration point L_1 of the Sun-Earth system at a position and with a velocity consistent with the HOI time.*

One can think of the TCM problem as a *rendezvous problem*. After launch, and the subsequent launch error, the actual spacecraft goes off the nominal trajectory. Since placement of the spacecraft on the nominal trajectory at the appropriate location *and time* is important for our problem, imagine that a virtual spacecraft remains on the nominal trajectory (corresponding to zero launch error). Our goal then is to perform maneuvers such that the actual spacecraft will rendezvous with the virtual spacecraft on the nominal trajectory. From the rendezvous point onward, the actual spacecraft will then be on the nominal trajectory.

The optimization problem as stated has two important features. First, it involves discontinuous controls, since the impulsive maneuvers are represented by jumps in the velocity of the spacecraft. A reformulation of the problem to cast it into the framework required by continuous optimal control algorithms will be discussed later in this section. Secondly, the final halo orbit insertion time T_{HOI}, as well as all intermediate maneuver times, must be included among the optimization parameters ($\mathbf{p}$). This too requires further reformulation of the dynamical model to capture the influence of these parameters on the solution at a given optimization iteration.

Next, we discuss the reformulations required to solve the HOI discontinuous control problem; modifications of the following procedure required to solve the MOI problem are discussed in §8.5. We assume that the evolution of the spacecraft is described by a generic set of six ordinary differential

equations (ODEs)

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}), \tag{8.3.1}$$

where $\mathbf{x} = (\mathbf{x}^p; \mathbf{x}^v) \in \mathbb{R}^6$ contains both positions ($\mathbf{x}^p$) and velocities ($\mathbf{x}^v$). The dynamical model of Equation (8.3.1) can be either the CR3BP or a more complex model that incorporates the influence of the Moon and other planets or non-gravitational effects. In this chapter, we use the CR3BP approximation to illustrate the method.

To deal with the discontinuous nature of the impulsive control maneuvers, the equations of motion (e.o.m.) are solved simultaneously on each interval between two maneuvers. Let the maneuvers take place at times T_i, $i = 1, 2, ..., n$ and let $\mathbf{x}_i(t)$, $t \in [T_{i-1}, T_i]$ be the solution of Equation (8.3.1) on the interval $[T_{i-1}, T_i]$ (see Figure 8.3.1).

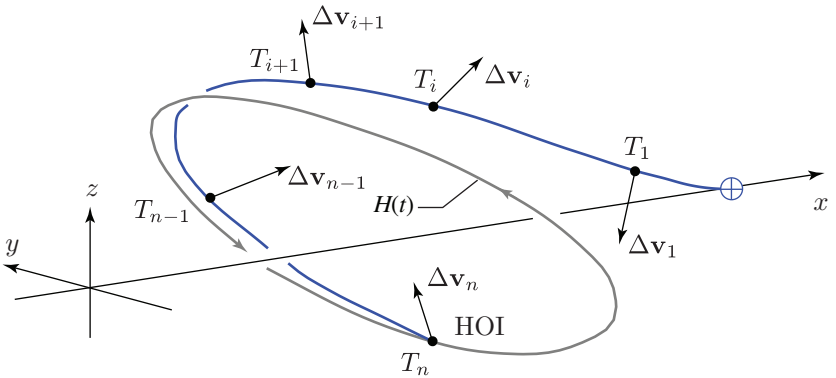


FIGURE 8.3.1. Transfer trajectory. Maneuvers take place at times $T_i, i = 1, 2, ..., n$. In the stable manifold insertion problem, there is no maneuver at T_n, i.e., $\Delta\mathbf{v}_n = \mathbf{0}$.

To capture the influence of the maneuver times on the solution of the e.o.m. and to be able to solve the e.o.m. simultaneously, we scale the time in each interval by the duration $\Delta T_i = T_i - T_{i-1}$. This is a standard technique for optimizing over the final integration time. As a consequence, all time derivatives in Equation (8.3.1) are scaled by $1/\Delta T_i$. The dimension of the dynamical system is thus increased to $N_x = 6n$.

Position continuity constraints are imposed at each maneuver, that is,

$$\mathbf{x}_i^p(T_i) = \mathbf{x}_{i+1}^p(T_i), \qquad i = 1, 2, ..., n-1. \tag{8.3.2}$$

In addition, the final position is forced to lie on the given halo orbit (or stable manifold) at the proper time, that is,

$$\mathbf{x}_n^p(T_n) = \mathbf{x}_H^p(T_n), \tag{8.3.3}$$

where the halo orbit is parameterized by the HOI time T_n. Recall that the halo orbit, $H(t) = (\mathbf{x}_H^p(t); \mathbf{x}_H^v(t))$, (and also the stable manifold $M(t)$),

is fixed and provided. Equation 8.3.3 is the constraint for the rendezvous problem at hand. If the insertion phase were not imposed, then the position on the halo orbit would be parameterized by an independent variable, i.e., the right hand side of Equation 8.3.3 would be $\mathbf{x}_H^p(\tau)$ where τ is a free optimization parameter.

Additional constraints dictate that the first maneuver (TCM1) is delayed by at least a prescribed amount $TCM1_{\min}$, that is,

$$T_1 \geq TCM1_{\min}, \tag{8.3.4}$$

and that the order of maneuvers is respected,

$$T_{i-1} < T_i < T_{i+1}, \qquad i = 1, 2, ..., n-1. \tag{8.3.5}$$

With a cost function defined as some measure of the velocity discontinuities

$$\begin{aligned} \Delta\mathbf{v}_i &= \mathbf{x}_{i+1}^v(T_i) - \mathbf{x}_i^v(T_i), \qquad i = 1, 2, ..., n-1, \\ \Delta\mathbf{v}_n &= \mathbf{x}_H^v(T_n) - \mathbf{x}_n^v(T_n), \end{aligned} \tag{8.3.6}$$

the optimization problem becomes

$$\min_{T_i, \mathbf{x}_i, \Delta\mathbf{v}_i} C(\Delta\mathbf{v}_i), \tag{8.3.7}$$

subject to the constraints in Equations (8.3.2)-(8.3.6). Note that velocities at the intermediate maneuver points and the final insertion point are matched using bursts $\Delta\mathbf{v}_i, i = 1, 2, ..., n-1$, and $\mathbf{v}_n$, respectively. More details on selecting the form of the cost function are given in §8.4.

Launch Errors and Sensitivity Analysis. In many optimal control problems, obtaining an optimal solution is not the only goal. The influence of problem parameters on the optimal solution (the so called sensitivity of the optimal solution) is also needed. Sensitivity information provides a first-order approximation to the behavior of the optimal solution when parameters are not at their optimal values or when constraints are slightly violated.

In the problems treated in this chapter, for example, we are interested in estimating the changes in fuel efficiency (ΔV) caused by possible perturbations in the launch velocity (ϵ_0^v) and by different delays in the first maneuver (TCM1). As we show in §8.4, the cost function is very close to being linear in these parameters ($TCM1_{\min}$ and ϵ_0^v). Therefore, evaluating the sensitivity of the optimal cost is a very inexpensive and accurate method of assessing the influence of different parameters on the optimal trajectory (especially in our problem).

In COOPT, we make use of the Sensitivity Theorem (see Bertsekas [1995]) for nonlinear programming problems with equality and/or inequality constraints:

8.3.1 Theorem. *Let f, h, and g be twice continuously differentiable and consider the family of problems*

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } h(x) = u, \quad g(x) \leq v, \end{aligned} \tag{8.3.8}$$

parameterized by the vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^r$. Assume that for $(u, v) = (0, 0)$ this problem has a local minimum x^*, which is regular and which together with its associated Lagrange multiplier vectors λ^* and μ^*, satisfies the second order sufficiency conditions. Then there exists an open sphere S centered at $(u, v) = (0, 0)$ such that for every $(u, v) \in S$ there is an $x(u, v) \in \mathbb{R}^n$, $\lambda(u, v) \in \mathbb{R}^m$, and $\mu(u, v) \in \mathbb{R}^r$, which are a local minimum and associated Lagrange multipliers of problem (8.3.8). Furthermore, $x(\cdot)$, $\lambda(\cdot)$, and $\mu(\cdot)$ are continuously differentiable in S and we have $x(0, 0) = x^*, \lambda(0, 0) = \lambda^*, \mu(0, 0) = \mu^*$. In addition, for all $(u, v) \in S$, We have

$$\begin{aligned} \nabla_u p(u, v) &= -\lambda(u, v), \\ \nabla_v p(u, v) &= -\mu(u, v), \end{aligned} \tag{8.3.9}$$

where $p(u, v)$ is the optimal cost parameterized by (u, v),

$$p(u, v) = f(x(u, v)). \tag{8.3.10}$$

The influence of delaying the maneuver TCM1 is thus directly computed from the Lagrange multiplier associated with the constraint of Equation (8.3.4). To evaluate sensitivities of the cost function with respect to perturbations in the launch velocity (ϵ_0^v), we must include this perturbation explicitly as an optimization parameter and fix it to some prescribed value through an equality constraint. That is, the launch velocity is set to

$$\mathbf{v}(0) = \mathbf{v}_0^{\text{nom}} \left(1 + \frac{\epsilon_0^v}{\|\mathbf{v}_0^{\text{nom}}\|} \right), \tag{8.3.11}$$

where $\mathbf{v}_0^{\text{nom}}$ is the nominal launch velocity and

$$\epsilon_0^v = \epsilon, \tag{8.3.12}$$

for a given ϵ. The Lagrange multiplier associated with the constraint in Equation (8.3.12) yields the desired sensitivity.

Description of the COOPT Software. COOPT is a software package for optimal control and optimization of systems modeled by differential-algebraic equations (DAE), developed by the Computational Science and Engineering Group at the University of California, Santa Barbara. It has been designed to control and optimize a general class of DAE systems, which may be quite large. Here we describe the basic methods used in COOPT. We consider the DAE system

$$\begin{aligned} \mathbf{F}(t, \mathbf{x}, \dot{\mathbf{x}}, \mathbf{p}, \mathbf{u}(t)) &= \mathbf{0}, \\ \mathbf{x}(t_1, \mathbf{r}) &= \mathbf{x}_1(\mathbf{r}), \end{aligned} \tag{8.3.13}$$

where the DAE is index zero, one, or semi-explicit index two (see Ascher and Petzold [1998], or Brenan, Campbell, and Petzold [1995]) and the initial conditions have been chosen so that they are consistent (that is, the constraints of the DAE are satisfied). The control parameters $\mathbf{p}$ and $\mathbf{r}$ and the vector-valued control function $\mathbf{u}(t)$ must be determined such that the objective function

$$\int_{t_1}^{t_{\max}} \Psi(t, \mathbf{x}(t), \mathbf{p}, \mathbf{u}(t))\, dt + \Theta(t_{\max}, \mathbf{x}(t_{\max}), \mathbf{p}, \mathbf{r}), \tag{8.3.14}$$

is minimized and some additional equality and/or inequality constraints

$$\mathbf{g}(t, \mathbf{x}(t), \mathbf{p}, \mathbf{r}, \mathbf{u}(t)) \gtreqless 0, \tag{8.3.15}$$

are satisfied. The optimal control function $\mathbf{u}^*(t)$ is assumed to be continuous. To represent $\mathbf{u}(t)$ in a low-dimensional vector space, we use piecewise polynomials on $[t_1, t_{\max}]$, where their coefficients are determined by the optimization. For ease of presentation we can therefore assume that the vector $\mathbf{p}$ contains both the parameters and these coefficients (we let N_p denote the combined number of these values) and discard the control function $\mathbf{u}(t)$ in the remainder of this section. Also, we consider that the initial states are fixed and therefore discard the parameters $\mathbf{r}$ from the formulation of the optimal control problem. Hence, we consider

$$\mathbf{F}(t, \mathbf{x}, \dot{\mathbf{x}}, \mathbf{p}) = 0, \qquad \mathbf{x}(t_1) = \mathbf{x}_1, \tag{8.3.16a}$$

$$\int_{t_1}^{t_{\max}} \psi(t, \mathbf{x}(t), \mathbf{p})\, dt + \Theta(t_{\max}, \mathbf{x}(t_{\max}), \mathbf{p}) \quad \text{minimized}, \tag{8.3.16b}$$

$$\mathbf{g}(t, \mathbf{x}(t), \mathbf{p}) \gtreqless 0. \tag{8.3.16c}$$

There are a number of well-known methods for direct discretization of the optimal control problem in Equations (8.3.16), for the case in which the DAEs can be reduced to ODEs in standard form. COOPT implements the *single shooting* method and a modified version of the *multiple shooting* method, both of which allow the use of adaptive DAE software.

In the multiple shooting method, the time interval $[t_1, t_{\max}]$ is divided into subintervals $[t_i, t_{i+1}]$ $(i = 1, \ldots, N_{tx})$, and the differential equations in Equation (8.3.16a) are solved over each subinterval, where additional intermediate variables $\mathbf{X}_i$ are introduced. On each subinterval we denote the solution at time t of Equation (8.3.16a) with initial value $\mathbf{X}_i$ at t_i by $\mathbf{x}(t, t_i, \mathbf{X}_i, \mathbf{p})$.

Continuity between subintervals in the multiple shooting method can be achieved via the continuity constraints

$$\mathbf{C}_1^i(\mathbf{X}_{i+1}, \mathbf{X}_i, \mathbf{p}) \equiv \mathbf{X}_{i+1} - \mathbf{x}(t_{i+1}, t_i, \mathbf{X}_i, \mathbf{p}) = \mathbf{0}. \tag{8.3.17}$$

The additional constraints of Equation (8.3.16c) are required to be satisfied at the boundaries of the shooting intervals

$$\mathbf{C}_2^i(\mathbf{X}_i, \mathbf{p}) \equiv \mathbf{g}(t_i, \mathbf{X}_i, \mathbf{p}) \gtreqless \mathbf{0}. \tag{8.3.18}$$

Following common practice, we write

$$\Phi(t) = \int_{t_1}^{t} \psi(\tau, \mathbf{x}(\tau), \mathbf{p})\, d\tau, \tag{8.3.19}$$

which satisfies $\Phi'(t) = \psi(t, \mathbf{x}(t), \mathbf{p})$, $\Phi(t_1) = 0$. This introduces another equation and variable into the differential system in Equation (8.3.16a). The discretized optimal control problem becomes

$$\min_{\mathbf{X}_2, \ldots, \mathbf{X}_{Ntx}, \mathbf{p}} \Phi(t_{\max}) + \Theta(t_{\max}), \tag{8.3.20}$$

subject to the constraints

$$\mathbf{C}_1^i(\mathbf{X}_{i+1}, \mathbf{X}_i, \mathbf{p}) = \mathbf{0}, \tag{8.3.21a}$$

$$\mathbf{C}_2^i(\mathbf{X}_i, \mathbf{p}) \gtreqless \mathbf{0}. \tag{8.3.21b}$$

This problem can be solved by an optimization algorithm. We use the solver SNOPT (see Gill, Murray, and Saunders [1997, 1998]), which incorporates a sequential quadratic programming (SQP) method (see Gill, Murray, and Wright [1981]). The SQP methods require a gradient and Jacobian matrix that are the derivatives of the objective function and constraints with respect to the optimization variables. We compute these derivatives via DAE sensitivity software DAE (Li and Petzold [2000]). The sensitivity equations to be solved by DAE are generated via the automatic differentiation software ADIFOR (see Bischof, Carle, Corliss, Griewank, and Hovland [1992] and Bischof, Carle, Hovland, Khademi, and Mauer [1998]).

This basic multiple shooting type of strategy can work very well for small-to-moderate size ODE systems, and has an additional advantage that it is inherently parallel. However, for large-scale ODE and DAE systems there is a problem because the computational complexity grows rapidly with the dimension of the ODE system. COOPT implements a highly efficient modified multiple shooting method (Gill, Jay, Leonard, Petzold, and Sharma [2000]; Serban and Petzold [2001]) which reduces the computational complexity to that of single shooting for large-scale problems. However, we have found it sufficient to use single shooting for the trajectory design problems treated in this chapter.

8.4 Numerical Results for the Halo Orbit Insertion Problem

Circular Restricted Three-Body Problem. As mentioned earlier, we use the equations of motion derived under the CR3BP assumption as the underlying dynamical model in Equation (8.3.1). In this model, it is assumed that the primaries (Earth and Sun in our case) move on circular orbits around the center of mass of the system and that the third body (the spacecraft) does not influence the motion of the primaries. We write the equations in a rotating frame, as in Figure 2.3.1.

Using nondimensional units, the equations of motion in the CR3BP model (cf. (2.3.8)), written as a set of first order ODEs, are

$$\begin{aligned}
\dot{x}_1 &= x_4 \\
\dot{x}_2 &= x_5 \\
\dot{x}_3 &= x_6 \\
\dot{x}_4 &= 2x_2 - \frac{\partial \bar{U}}{\partial x_1} \\
\dot{x}_5 &= -2x_1 - \frac{\partial \bar{U}}{\partial x_2} \\
\dot{x}_6 &= \frac{\partial \bar{U}}{\partial x_3}
\end{aligned} \tag{8.4.1}$$

where

$$\begin{aligned}
\mathbf{x} &= [x_1, x_2, x_3, x_4, x_5, x_6]^T = [x, y, z, v_x, v_y, v_z]^T \\
\bar{U} &= -\frac{1}{2}(x_1^2 + x_2^2) - \frac{1-\mu}{r_1} - \frac{\mu}{r_2} - \frac{1}{2}\mu(1-\mu) \\
r_1 &= \left((x_1+\mu)^2 + x_2^2 + x_3^2\right)^{1/2} \\
r_2 &= \left((x_1-1+\mu)^2 + x_2^2 + x_3^2\right)^{1/2}
\end{aligned} \tag{8.4.2}$$

and μ is the ratio between the mass of the Earth and the mass of the Sun-Earth system,

$$\mu = \frac{m_\oplus}{m_\oplus + m_\odot}, \tag{8.4.3}$$

where $\oplus$ denotes the Earth and $\odot$ the Sun. For the Sun-Earth system, we use $\mu = 3.03591 \times 10^{-6}$. In the above equations, time is scaled by the period of the primaries orbits ($T/2\pi$, where $T = 1$ year), positions are scaled by the Sun-Earth distance ($L \equiv d_{\oplus\odot} = 1.49597927 \times 10^8$km), and velocities are scaled by the Earth's average orbital speed around the Sun ($2\pi L/T = 29.80567$ km/s). We note that the values used are more precise than the values for this system given in Table 2.2.1.

Choice of Cost Function. At this point we need to give some more details on the choice of an appropriate cost function for the optimization problem (8.3.7). Typically in space missions, the spacecraft performance is

measured in terms of the maneuver sizes $\Delta \mathbf{v}_i$. We consider the following two cost functions.

$$C_1(\Delta \mathbf{v}) = \sum_{i=1}^{n} \|\Delta \mathbf{v}_i\|^2 \tag{8.4.4}$$

and

$$C_2(\Delta \mathbf{v}) = \sum_{i=1}^{n} \|\Delta \mathbf{v}_i\| \tag{8.4.5}$$

where $\|\cdot\|$ denotes the usual Euclidean norm.

While the second of these may seem physically the most meaningful, as it measures the total sum of the maneuver sizes, such a cost function is nondifferentiable whenever one of the maneuvers vanishes. In our case, this problem occurs already at the first optimization iteration, as the initial guess transfer trajectory only has a single nonzero maneuver at halo insertion. The first cost function, on the other hand, is differentiable everywhere.

Although the cost function C_1 is more appropriate for the optimizer, it raises two new problems. Not only is it not as physically meaningful as the cost function C_2, but, in some particular cases, decreasing C_1 may actually lead to increases in C_2.

To resolve these issues, we use the following three-stage optimization sequence:

1. Starting with the nominal transfer trajectory as initial guess, and allowing initially n maneuvers, we minimize C_1 to obtain a first optimal trajectory, $\mathcal{T}_1^*$.

2. Using $\mathcal{T}_1^*$ as initial guess, we minimize C_2 to obtain $\mathcal{T}_2^*$. It is possible that during this optimization stage some maneuvers can become very small. After each optimization iteration we monitor the feasibility of the iterate and the sizes of all maneuvers. As soon as at least one maneuver decreases under a prescribed threshold (typically 0.1 m/s) at some feasible configuration, we stop the optimization algorithm.

3. If necessary, a third optimization stage, using $\mathcal{T}_2^*$ as initial guess and C_2 as cost function is performed with a reduced number of maneuvers $\bar{n}$ (obtained by removing those maneuvers identified as "zero maneuvers" in step 2).

Merging Optimal Control with Dynamical Systems Theory. Next, we present results for the halo orbit insertion problem and later for the stable manifold insertion problem (§8.5). In both cases we are investigating the effect of varying times for $TCM1_{\min}$ on the optimal trajectory, for given perturbations in the nominal launch velocity. The staggered optimization procedure described above is applied for values of $TCM1_{\min}$ ranging from 1 day to 5 days and perturbations in the magnitude of the launch velocity

ϵ_0^v ranging from -7 m/s to $+7$ m/s. We present typical transfer trajectories, as well as the dependency of the optimal cost on the two parameters of interest. In addition, using the algorithm presented in §8.3, we perform a sensitivity analysis of the optimal solution. For the Genesis TCM problem it turns out that sensitivity information of first order is sufficient to characterize the influence of $TCM1_{\min}$ and ϵ_0^v on the spacecraft performance.

The merging of optimal control and dynamical systems theory has been done through the use of invariant manifolds, in this case the stable invariant manifold of the target halo orbit. This merging can take one of two forms: (1) the use of a nominal transfer trajectory which is near the invariant manifold as an accurate initial guess, or (2) the targeting of the stable invariant manifold itself.

Halo Orbit Insertion (HOI) Problem. In this problem we directly target the selected halo orbit with the last maneuver taking place at the HOI point. Using the optimization procedure described in the previous section, we compute the optimal cost transfer trajectories for various combinations of $TCM1_{\min}$ and ϵ_0^v. In all of our computations, the launch conditions are those corresponding to the nominal transfer trajectory, i.e.,

$$\begin{aligned}
x_0^{\text{nom}} &= 1.496032475412839 \times 10^8 \text{ km} \\
y_0^{\text{nom}} &= 1.943203061350240 \times 10^3 \text{ km} \\
z_0^{\text{nom}} &= -2.479095822700627 \times 10^3 \text{ km} \\
(v_0^{\text{nom}})_x &= -4.612683390613825 \text{ km/s} \\
(v_0^{\text{nom}})_y &= 9.412034579485869 \text{ km/s} \\
(v_0^{\text{nom}})_z &= -3.479627336419212 \text{ km/s}
\end{aligned}$$

with the launch velocity perturbed as described in §8.3. These initial conditions are given in the Earth-Sun barycentered rotating frame.

As an example, we present complete results for the case in which the launch velocity is perturbed by -3 m/s and the first maneuver correction is delayed by at least 3 days. Initially, we allow for $n = 4$ maneuvers. In the first optimization stage, the second type of cost function has a value of $C_1^* = 1153.998$ $(\text{m/s})^2$ after 5 iterations. This corresponds to $C_2^* = 50.9123$ m/s. During the second optimization stage, we monitor the sizes of all four maneuvers, while minimizing the cost function C_1. After 23 iterations, the optimization was interrupted at a feasible configuration when at least one maneuver decreased below a preset tolerance of 0.1 m/s. The corresponding cost function is $C_2^{**} = 45.1216$ m/s with four maneuvers of sizes 33.8252 m/s, 0.0012 m/s, 0.0003 m/s, and 11.2949 m/s. In the last optimization stage we remove the second and third maneuvers and again minimize the cost function C_2. After 7 optimization iterations an optimal solution with $C_2^{***} = 45.0292$ m/s is obtained. The two maneuvers of the optimal trajectory have sizes of 33.7002 m/s and 11.3289 m/s and take

place at 3.0000 and 110.7969 days after launch, respectively. The resulting optimal trajectory is presented in Figure 8.4.1.

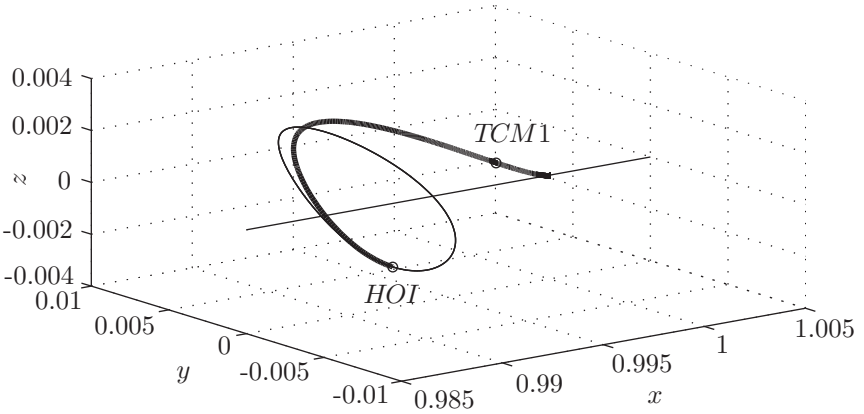


FIGURE 8.4.1. HOI problem. Optimal transfer trajectory for $TCM1_{\min} = 4$ days, $\epsilon_0^v = 3$ m/s, and $n = 4$. The optimal trajectory has $\bar{n} = 2$ maneuvers (represented by circles).

Lagrange multipliers associated with the constraints of Equations (8.3.4) and (8.3.12) give the sensitivities of the optimal solution with respect to launching velocity perturbation, −10.7341 (m/s)/(m/s), and delay in the first maneuver correction, 4.8231 (m/s)/day.

Computational and Communication Times. All experiments were performed on a PC workstation with an Intel Pentium III 800 MHz processor running Linux 2.2.12. The code was compiled with gcc with second level optimization. A typical run (the first optimization stage of the case presented in Figure 8.4.1) takes 0.29 s for the problem set-up and 4.23 s for the actual optimization.

For an actual TCM, this computation step is practially instantaneous. Most of the time delay is taken up before this step, by the determination of the position and velocity of the spacecraft, which reveals the launch error. Once the appropriate maneuver has been computed on the ground, it takes only a matter of seconds to communicate the maneuver information to the spacecraft.

Accuracy. We want to compute maneuvers at least as accurately as they can be implemented, given the accuracy to which the spacecraft's position and velocity can be measured. For a mission in the Earth's neighborhood, such as Genesis, the measurement accuracy is about 1 km for position and about 0.01 m/s for velocity. For all of our computations, the integration accuracy was well within these measurement limits.

Launch Errors and Sensitivity Analysis. The staggered optimization procedure is applied for all values of $TCM1_{\min}$ and ϵ_0^v in the regime of interest. In a first experiment, we investigate the possibility of correcting for errors in the launch velocity using at most two maneuvers ($n = 2$). The surface of optimal cost (C_2 in m/s) as a function of these two parameters is presented in Figure 8.4.2. Numerical values are given in Table 8.4.1.

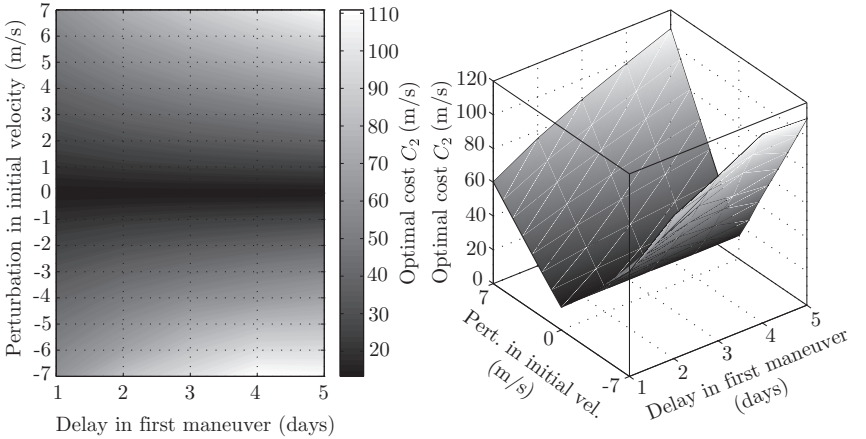


FIGURE 8.4.2. HOI problem. Influence of $TCM1_{\min}$ and ϵ_0^v on the optimal cost (C_2 in m/s) for $n = 2$.

Except for the cases in which there is no error in the launch velocity (and for which the final optimal transfer trajectories have only one maneuver at HOI), the first correction maneuver is always on the prescribed lower bound $TCM1_{\min}$. The evolution of the time at which the halo insertion maneuver takes place as a function of the two parameters considered is shown in Figure 8.4.3.

Recalling that the nominal transfer trajectory has $T_{\text{HOI}} = 110.2$ days, it follows that, for all cases investigated, halo orbit insertion takes place at most 18.6 days earlier or 28.3 days later than in the nominal case.

Several important observations can be drawn from these results. First, it can be seen that, for all cases that we investigated, the optimal costs are well within the ΔV budget allocated for transfer correction maneuvers (150 m/s for the Genesis mission). Secondly, as the second plot in Figure 8.4.2 shows, the cost function surface is very close to being linear with respect to both $TCM1_{\min}$ time and launch velocity error. This suggests that first order derivative information, as obtained from sensitivity analysis of the optimal solution (§8.3), provides a very good approximation to the surface.

TABLE 8.4.1. HOI problem. Optimal costs (C_2 in m/s) for different launch velocity perturbations and delays in first trajectory correction maneuver for $n = 2$.

ϵ_0^v	TCM1 (days)				
(m/s)	1	2	3	4	5
-7	64.8086	76.0845	88.4296	99.6005	109.9305
-6	54.0461	67.0226	77.7832	86.8630	95.8202
-5	47.1839	57.9451	66.6277	74.4544	81.8284
-4	40.2710	48.8619	55.8274	62.0412	67.9439
-3	33.4476	39.8919	45.0290	49.6804	54.1350
-2	26.6811	30.9617	34.3489	37.3922	40.3945
-1	19.9881	22.2715	23.7848	25.2468	26.6662
0	13.4831	13.3530	13.4606	13.3465	13.2919
1	23.1900	21.9242	23.2003	24.4154	25.5136
2	26.2928	30.2773	33.3203	35.9203	38.3337
3	34.6338	38.8496	43.5486	47.7200	51.6085
4	41.4230	47.5266	53.9557	62.3780	65.1411
5	45.9268	56.2245	64.4292	75.0188	81.4325
6	53.9004	64.9741	76.6978	83.8795	95.2313
7	61.4084	75.9169	85.4875	98.4197	106.0411

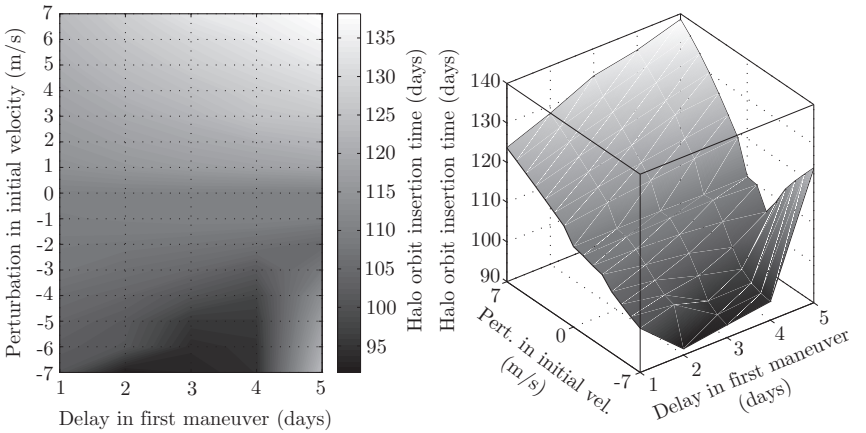


FIGURE 8.4.3. HOI problem. Influence of $TCM1_{\min}$ and ϵ_0^v on the halo orbit insertion time (T_{HOI} in days) for $n = 2$.

For a few points on the cost function surface, we present tangents obtained from sensitivity data in Figure 8.4.4.

Finally, the halo orbit insertion time is always close enough to that of the nominal trajectory so as to not affect either the collection of the solar wind or the rest of the mission (mainly the duration for which the spacecraft evolves on the halo orbit before initiation of the return trajectory).

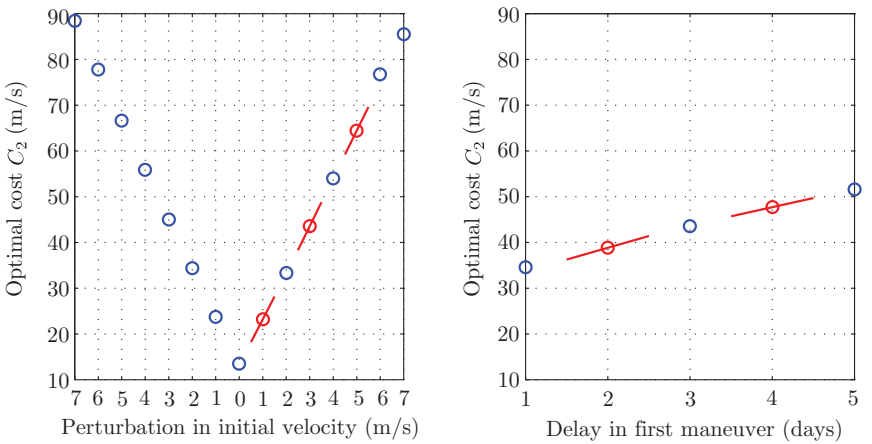


FIGURE 8.4.4. HOI problem. Sensitivity of the optimal solution for $n = 2$. Circles correspond to optimization results. Line segments are predictions based on sensitivity computations. The figure on the left was obtained with $TCM1_{\min} = 3$ days and shows the sensitivity of the optimal solution with respect to ϵ_0^v. The figure on the right was obtained with $\epsilon_0^v = 3$ m/s and shows the sensitivity of the optimal solution with respect to $TCM1_{\min}$.

In a second set of numerical experiments, we allow initially for as many as $n = 4$ maneuvers. This additional degree of freedom in the optimization leads to further reductions in the optimal cost function, as data in Table 8.4.2 shows.

The corresponding cost function surface is presented in Figure 8.4.5.

It is interesting to note that all optimal transfer trajectories have $\bar{n} = 2$ maneuvers for negative errors in the launch velocity, $\bar{n} = 1$ maneuver if there is no error, and $\bar{n} = 3$ maneuvers for positive launch velocity errors. As in the previous case, the time for the first correction maneuver is always on the prescribed lower bound (i.e., $TCM1 = TCM1_{\min}$), while the halo orbit insertion time, shown in Figure 8.4.6, is at most 2.6 days earlier or 21.4 days later than in the nominal case.

8.5 Numerical Results for the Stable Manifold Orbit Insertion Problem

Obtaining a Good Initial Guess. In the MOI problem the last nonzero maneuver takes place on the stable manifold and there is no maneuver to insert onto the halo orbit. This implies that, in addition to the constraints of Equation (8.3.3) imposing that the final position is on the halo orbit,

TABLE 8.4.2. HOI problem. Optimal costs (C_2 in m/s) for different launch velocity perturbations and delays in first transfer correction maneuver for the best case over $n = 2, 3, 4$.

ϵ_0^v	TCM1 (days)				
(m/s)	1	2	3	4	5
-7	61.0946	76.0852	88.4295	99.3123	109.9174
-6	54.0461	67.0212	77.7832	86.8994	95.8202
-5	47.1389	57.9277	66.6277	74.4513	81.8572
-4	40.2710	48.8619	55.7984	62.0398	67.9438
-3	33.3664	39.8919	45.0290	49.6804	54.1357
-2	26.6720	30.9617	34.3489	37.3911	40.3945
-1	19.9674	22.1091	23.7848	25.2640	26.6618
0	13.4598	13.2902	13.4428	13.2907	13.2919
1	19.8257	21.9026	23.2005	24.4149	25.4359
2	26.2933	30.2773	33.3077	35.9203	38.3337
3	32.8151	38.8496	43.5486	47.7200	51.6085
4	39.3646	47.5279	53.9557	59.7078	65.1117
5	45.9127	56.2333	64.4292	71.7790	78.7022
6	52.4968	64.9741	74.9477	83.8795	92.3090
7	59.0967	73.7398	85.4875	95.9822	105.8960

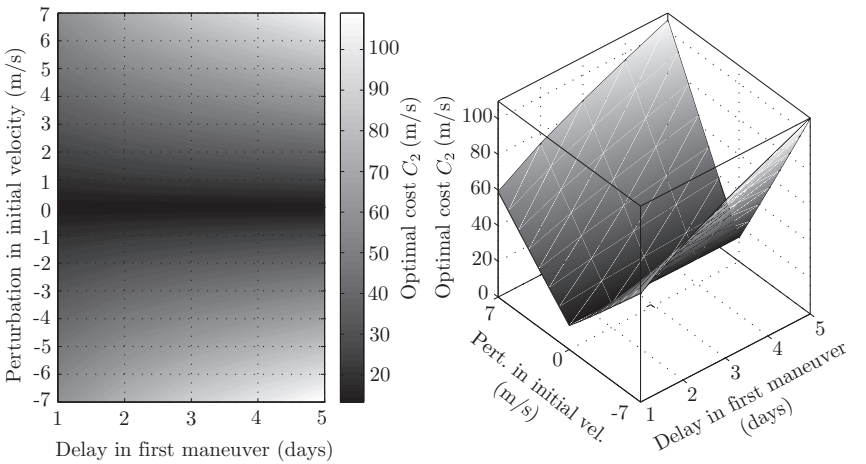


FIGURE 8.4.5. HOI problem. Influence of $TCM1_{\min}$ and ϵ_0^v on the optimal cost (C_2 in m/s). In each case, the best trajectory over $n = 2, 3, 4$ was plotted.

constraints must be imposed to match the final spacecraft velocity with the velocity on the halo orbit. These highly nonlinear constraints, together with the fact that a much larger parameter space is now investigated (we target an entire surface as opposed to just a curve) make the optimization

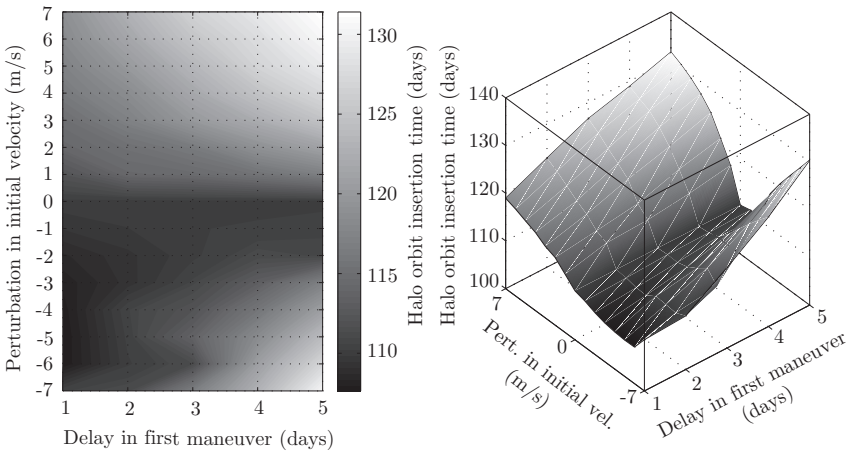


FIGURE 8.4.6. HOI problem. Influence of $TCM1_{\min}$ and ϵ_0^v on the halo orbit insertion time (T_{HOI} in days). In each case, the best trajectory over $n = 2, 3, 4$ was plotted.

problem much more difficult than the one corresponding to the HOI case. The first problem that arises is that the nominal transfer trajectory is not a good enough initial guess to ensure convergence to an optimum. To obtain an appropriate initial guess we use the following procedure:

Step 1. We start by selecting an HOI time, T_{HOI}. This yields the position and velocity on the halo orbit.

Step 2. The above position and velocity are perturbed in the direction of the stable manifold and the equations of motion in Equation (8.4.1) are then integrated backwards in time for a selected duration T_S. This yields an MOI point which is now fixed in time, position, and velocity.

Step 3. For a given value of $TCM1_{\min}$ and with $\epsilon_0^v = 0$, and using the nominal transfer trajectory as initial guess, we use COOPT to find a trajectory that targets this MOI point, while minimizing C_1.

With the resulting trajectory as an initial guess and the desired value of ϵ_0^v we proceed with the staggered optimization presented before to obtain the final optimal trajectory for insertion on the stable manifold. During the three stages of the optimization procedure, both the MOI point and the HOI point are free to move (in position, velocity, and time) on the stable manifold surface and on the halo orbit, respectively.

The fact that we are using local optimization techniques implies that the computed optimal trajectories are very sensitive to the choice of the initial guess trajectory. For given values of the problem parameters (such as initial number of maneuvers, perturbation in launch velocity, and lower

bound on TCM1) we find optimal trajectories in a neighborhood of the initial guess trajectory. In other words, computed optimal trajectories can be 'steered' towards regions of interest by appropriate choices of initial guess trajectories. For example, taking the launch time to be $T_L = 0$ and the HOI time (T^*_{HOI}) of the nominal transfer trajectory as a reference point on the halo orbit, we can investigate a given zone of the design space by an appropriate choice of the HOI point of our initial guess trajectory with respect to T^*_{HOI} (step 1 of the above procedure). That is, we select a value T_0 such that $T_{\text{HOI}} = T^*_{\text{HOI}} + T_0$. The point where the initial guess trajectory inserts onto the stable manifold is then defined by selecting the duration T_S for which the equations of motion are integrated backwards in time (step 2 of the above procedure). This gives a stable manifold insertion time of $T_{\text{MOI}} = T_{\text{HOI}} - T_S = T^*_{\text{HOI}} + T_0 - T_S$. Next, we use COOPT to evaluate these various choices for the initial guess trajectories (step 3 of the above procedure). A schematic representation of this procedure is shown in Figure 8.5.1.

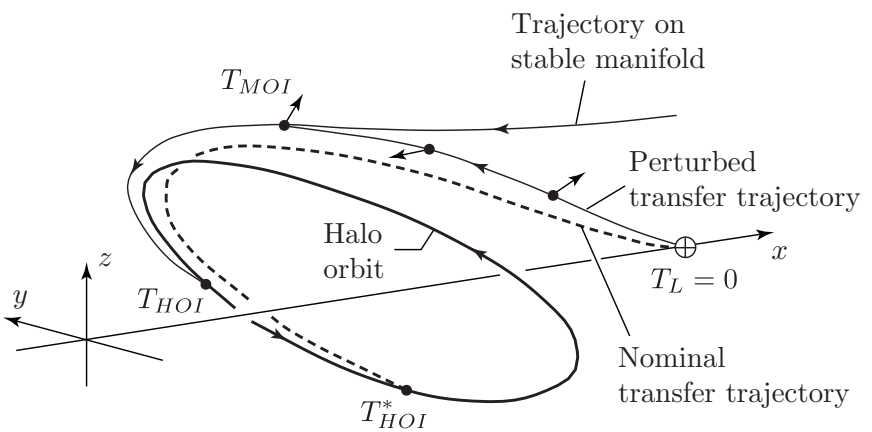


FIGURE 8.5.1. MOI Problem. Description of the initial guess computation procedure.

For different combinations of T_0 and T_S, Table 8.5.1 presents values of $C_2^*(\Delta\mathbf{v}) = \sum_{i=1}^{n} \|\Delta\mathbf{v}_i\|$ corresponding to the optimal initial guess trajectory that targets the resulting MOI point. Note that, for a given value T_0, there exists a value T_S for which we are unable to compute an initial guess trajectory. This is due to the fact that, for these values of T_0 and T_S, the resulting T_{MOI} is too small for COOPT to find a trajectory that targets the MOI point from $T_L = 0$.

Regions Best Suited for MOI Insertion. From the data given in Table 8.5.1 we can identify regions of the stable manifold that are best suited for MOI insertion. Examples of such regions are:

TABLE 8.5.1. MOI problem. Initial guess trajectories obtained for different choices of the parameters T_0 and T_S. All times are given in nondimensional units.

T_0	T_{HOI}	T_S	T_{MOI}	C_2^* (m/s)
-0.25	1.65916	0.25	1.40916	45.7317
		0.50	1.15916	93.2419
		0.75	-	-
0.00	1.90916	0.25	1.65916	21.7515
		0.50	1.40916	45.1291
		0.75	1.15916	94.0839
		1.00	-	-
0.75	2.65916	0.25	2.40916	21.1051
		0.50	2.15916	21.4791
		0.75	1.90916	24.8072
		1.00	1.65916	23.9035
		1.25	1.40916	43.2514
		1.50	1.15916	86.1323
		1.75	-	-
1.50	3.40916	0.25	3.15916	15.9145
		0.50	2.90916	16.2152
		0.75	2.65916	15.6983
		1.00	2.40916	17.6370
		1.25	2.15916	27.5903
		1.50	1.90916	18.9711
		1.75	1.65916	19.4283
		2.00	1.40916	28.3686
		2.25	1.15916	51.8521
		2.50	0.90916	105.7831
		2.75	0.65916	212.9997
		3.00	0.40916	519.7044
		3.25	-	-

- (*Region A*) MOI trajectories that insert to the halo orbit in the same region as the nominal transfer trajectory and which therefore correspond to initial guess trajectories with small T_0;
- (*Region B*) MOI trajectories that have HOI points on the "far side" of the halo orbit and which correspond to initial guess trajectories with halo insertion time around $T^*_{\text{HOI}} + 1.50$ ($T_0 = 1.50 \cdot 365/2\pi = 174.27$ days).

These choices are confirmed by the examples from Wilson, Howell, and Lo [1999]. Trajectories in Region B might, at first glance, appear unsuited for the Genesis mission as they would drastically decrease the duration for which the spacecraft evolves on the halo orbit (recall that design of

the return trajectory dictates the time at which the spacecraft must leave the halo orbit). However, as the typical MOI trajectory of Figure 8.5.2 shows, all trajectories on the stable manifold asymptotically wind onto the halo orbit and are thus very close to the halo orbit for a significant time. This means that collection of solar wind samples can start much earlier than halo orbit insertion, therefore providing enough time for all scientific experiments before the spacecraft leaves the halo orbit.

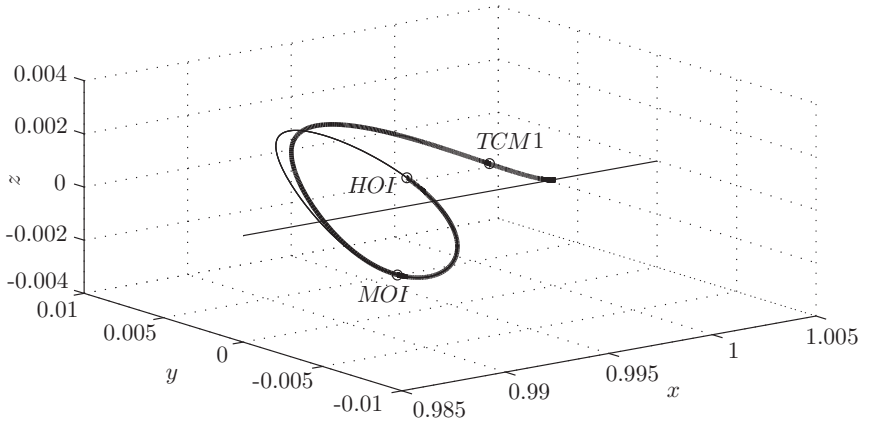


FIGURE 8.5.2. MOI Problem. Optimal transfer trajectory for $TCM1_{\min} = 4$ days, $\epsilon_0^v = -3$ m/s, and $n = 4$. The optimal trajectory has a cost function of $C_2 = 49.1817$ m/s and $\bar{n} = 2$ maneuvers. The first maneuver takes place at $TCM1 = 4$ days, the second one at $T_{\text{MOI}} = 112.11$ days, while HOI takes place $T_{\text{HOI}} = 173.25$ days after launch.

Once we select a region of the stable manifold by selecting an appropriate initial guess trajectory, we can perform the same type of analysis as done for the HOI problem of §8.4. In what follows, we consider the case in which we correct for perturbations in launch velocity by seeking optimal MOI trajectories in Region B, that is, on the far side of the halo from the Earth. For given values of ϵ_0^v and $TCM1_{\min}$, we first compute an MOI initial guess trajectory with $T_0 = 1.50$ and $T_S = 0.75$ and then use the staggered optimization procedure described in §8.4 to find an optimal MOI trajectory in this vicinity.

We present results from such computations in Table 8.5.2. It can be seen that the optimal MOI trajectories are very close (in terms of their associated cost function C_2) to the corresponding HOI trajectories. These results can be understood if we recall that the nominal transfer trajectory that we use in our experiments actually inserts onto the halo orbit directly as opposed to the manifold. To take full advantage of the stable manifold in correcting for launching errors, one may need to start with a nominal transfer trajectory that inserts onto the stable manifold. For missions that are designed to have such nominal transfer trajectories, correction trajecto-

ries that also insert onto the stable manifold are expected to be much more efficient than those obtained with the current formulation of the problem.

TABLE 8.5.2. MOI problem. Optimal costs (C_2 in m/s) for different launching velocity perturbations and delays in first transfer correction maneuver.

$TCM1_{\min}$ (days)	ϵ_0^v (m/s)	C_2 (m/s)
3	-3	45.1427
	-4	55.6387
	-5	65.9416
	-6	76.7144
	-7	87.3777
4	-3	49.1817
	-4	61.5221
	-5	73.4862
	-6	85.7667
	-7	99.3405
5	-3	53.9072
	-4	66.8668
	-5	81.1679
	-6	94.3630
	-7	109.2151

Summary and Remarks. In §8.2-8.5, we explored new approaches for automated parametric studies of optimal transfer correction maneuvers for a halo orbit mission. Using the halo orbit insertion approach, for all the launch velocity errors and $TCM1_{\min}$ considered we found optimal recovery trajectories. The cost functions (fuel consumption in terms of ΔV) are within the allocated budget even in the worst case (largest $TCM1_{\min}$ and largest launch velocity error).

Using the stable manifold insertion approach, we obtained similar results to those found using HOI targeted trajectories. The failure of the MOI approach to reduce the ΔV significantly may be because the optimization procedure (even in the HOI targeted case) naturally finds trajectories 'near' the stable manifold.

The main contribution of dynamical systems theory to the problem of finding optimal recovery trajectories is in the construction of good initial guess trajectories in sensitive regions which allows the optimizer to home in on the solution. We feel that this aspect of our work will be important in many other future mission design problems. Many missions in the future will also require the use ofoptimal control in the context of low thrust. The software and methods of this chapter can be used with little change for such problems.

8.6 Station-Keeping with the Target Point Approach

Station-Keeping for Libration Point Orbits. Periodic and quasi-periodic orbits near libration points are regularly chosen as nominal orbits for missions. Because of the instability of these orbits, occasional "nudges" from on-board thrusters are necessary to keep the spacecraft on track–near the nominal orbit.

Once the nominal orbit is specified, strategies must be implemented to keep the actual spacecraft trajectory sufficiently close to the nominal path. As in the previous sections, in this and the next section all maneuvers are assumed to be impulsive and occur at discrete times. We outline two approaches: Target Point and Floquet Mode. In this section, we look at the Target Point approach.

Target Point Station-Keeping. The goal of the Target Point station-keeping algorithm is to compute and implement maneuvers to maintain a spacecraft "close" to the nominal orbit, i.e., within a tube of some specified radius centered about the reference path. To accomplish this task, a control procedure is derived from minimization of a cost function. The cost function C is defined by weighting both the control energy required to implement a station-keeping maneuver, $(\Delta\mathbf{v})$, and a series of expected deviations of the six-dimensional state from the nominal orbit at specified future times t_i. This procedure was developed by Howell and Pernicka [1988] and Howell and Gordon [1994].

The Cost Function. The cost function to be minimized is written as

$$C = \Delta\mathbf{v}^T Q \Delta\mathbf{v} + \sum_{i=1}^{n} \delta\mathbf{p}_i{}^T R_i^p \delta\mathbf{p}_i + \sum_{i=1}^{n} \delta\mathbf{v}_i{}^T R_i^v \delta\mathbf{v}_i, \tag{8.6.1}$$

where superscript T denotes transpose. The variables in the cost function include the corrective maneuver, $\Delta\mathbf{v}$ at some time t_c. The $\delta\mathbf{p}_i$ are defined as 3×1 column vectors representating linear approximations of the expected deviations of the actual spacecraft trajectory from the nominal path at specified future times t_i. Likewise, the 3×1 vectors $\delta\mathbf{v}_i$ represent deviations of the spacecraft velocity at the corresponding t_i. The future times at which predictions of the position and velocity state of the vehicle are compared to the nominal path are denoted as ***target points***. They are represented as Δt_i such that $t_i = t_0 + \Delta t_i$ where t_0 is the initial reference time. In the following study, the case of three future target points will be presented.

In equations (8.6.1), Q, R_i^p and R_i^v are 3×3 weighting matrices. The weighting matrix Q is symmetric positive definite; the other weighting matrices are symmetric positive semi-definite. The weighting matrices are generally treated as constants that must be specified as inputs. Selection of

appropriate weighting matrix elements is a trial and error process that has proven to be time-consuming. A methodology has been developed that automatically selects and updates the weighting matrices for each maneuver. This "time-varying" weighting matrix algorithm is based solely on empirical observations.

Computation of $\Delta\mathbf{v}$. For simplicity, let us assume that the maneuver is implemented at $t_c = t_0$ with $(\delta\mathbf{p}_0, \delta\mathbf{v}_0)$ as the initial position and velocity deviations. Then the state transition matrix provides the following relationship:

$$\begin{aligned}\begin{bmatrix} \delta\mathbf{p}_i \\ \delta\mathbf{v}_i \end{bmatrix} &= \Phi(t_i, t_0) \begin{bmatrix} \delta\mathbf{p}_0 \\ \delta\mathbf{v}_0 + \Delta\mathbf{v} \end{bmatrix} \\ &= \begin{bmatrix} A_{i0} & B_{i0} \\ C_{i0} & D_{i0} \end{bmatrix} \begin{bmatrix} \delta\mathbf{p}_0 \\ \delta\mathbf{v}_0 + \Delta\mathbf{v} \end{bmatrix}.\end{aligned} \tag{8.6.2}$$

This allows us to compute the expected deviations $(\delta\mathbf{p}_i, \delta\mathbf{v}_i)$ at the target time t_i in terms of initial deviations $(\delta\mathbf{p}_0, \delta\mathbf{v}_0)$ and the maneuver $\Delta\mathbf{v}$. Therefore, the cost function C can be written in terms of submatrices of the state transiton matrix and the variables $\delta\mathbf{p}_0, \delta\mathbf{v}_0$ and $\Delta\mathbf{v}$.

Determination of the $\Delta\mathbf{v}$ corresponding to the relative minimum of this cost function, denoted $\Delta\mathbf{v}^*$, shows that the required maneuver $\Delta\mathbf{v}$ depends linearly on the position and velocity deviations needed to be corrected:

$$\begin{aligned}\Delta\mathbf{v}^* = &- \left[Q + \sum_{i=1}^{3} B_{i0}^T R_i^p B_{i0} + D_{i0}^T R_i^v D_{i0} \right]^{-1} \\ &\times \left[\left(\sum_{i=1}^{3} B_{i0}^T R_i^p B_{i0} + D_{i0}^T R_i^v D_{i0} \right) \delta\mathbf{v}_0 + \left(\sum_{i=1}^{3} B_{i0}^T R_i^p A_{i0} + D_{i0}^T R_i^v C_{i0} \right) \delta\mathbf{p}_0 \right].\end{aligned}$$

The performance of the Target Point algorithm is not truly "optimal," though it has been flight demonstrated to successfully control spacecraft at reasonable costs. This accomplishment alone provides the mission design with a quick and efficient way to obtain reasonable station-keeping strategies. Given some procedure to select the weighting matrices, the maneuver is computed from the above equation. The corrective maneuver $(\Delta\mathbf{v})$ is a function of spacecraft drift (in both position and velocity with respect to the nominal orbit), the state transition matrix elements associated with the nominal orbit, and the weighting matrices. Note that this general method could certainly accommodate inclusion of additional target points. Although the nominal orbit that is under consideration here is periodic, the methodology does not rely on periodicity; it could be applied to any type of motion.

Additonal Constraints. Three additional constraints are specified in the station-keeping procedure to restrict maneuver implementation.

1. The time elapsed between successive maneuvers must be greater than or equal to a specified minimum time interval, $\Delta t_{\min}$. This constraint

may be regulated by the orbit determination process, scientific payload requirements, and mission operations. Time intervals of one to three days are considered for the Earth-Moon libration points and 40-80 days for Sun-Earth libration points.

2. The second constraint is a scalar distance ($d_{\min}$) that specifies a minimum deviation from the nominal path *in position space* that must be exceeded prior to maneuver execution. For distances less than $d_{\min}$, maneuver computations do not occur.

3. In the Target Point approach, the magnitude of position deviations are compared between successive tracking intervals. If the magnitude is decreasing, a maneuver is not calculated.

For a corrective maneuver to be computed, all three criteria must be satisfied simultaneously.

After a maneuver is calculated by the algorithm, an additional constraint is specified on the minimum allowable maneuver magnitude, $\Delta v_{\min}$. If the magnitude of the calculated $\Delta\mathbf{v}$ is less than $\Delta v_{\min}$, then the recommendated maneuver is cancelled. This constraint is useful in avoiding "small" maneuvers that are approximately the same order of magnitude as the maneuver errors. It also serves to model actual hardware limitations, as modern propulsion devices have restrictions concerning the minimal control impulse that can be accurately executed.

8.7 Station-Keeping with Floquet Mode Approach

An alternate strategy for station-keeping is the Floquet Mode approach, a method that is significantly different from the Target Point approach. As discussed in §7.2, the behavior of the flow near a nominal halo orbit can be determined by the eigenvalues $\lambda_i, i = 1, \ldots, 6$ and the eigenvectors $\mathbf{e}_i, i = 1, \ldots, 6$ of M, the monodromy matrix.

The emphasis for the Floquet Mode approach is on formulating a ***controller*** which will effectively eliminate the unstable component of the error vector

$$\delta\mathcal{E}(t) = (\delta x, \delta y, \delta z, \delta\dot{x}, \delta\dot{y}, \delta\dot{z}),$$

defined as the difference between the actual coordinates obtained by tracking and the nominal ones computed isochronously on the nominal orbit.

At any epoch, t, $\delta\mathcal{E}(t)$ can be expressed in terms of the ***Floquet modes***, $\bar{\mathbf{e}}_i(t)$, as

$$\delta\mathcal{E}(t) = \sum_{1}^{6} \alpha_i \bar{\mathbf{e}}_i(t).$$

The objective of the controller is to add a maneuver

$$\Delta\mathbf{v} = (0, 0, 0, \Delta v_x, \Delta v_y, \Delta v_z)$$

such that *the magnitude of the component of the error vector in the unstable direction, α_1, is reduced to zero.* The five remaining components do not produce large departures from the nominal orbit. By contrast, the component of the error vector along the unstable mode increases by a factor of λ_1 ($\gg 1$) in each revolution.

Note that it is convenient to use the Floquet modes $\bar{\mathbf{e}}_i(t)$ instead of the vector $\mathbf{e}_i(t)$ to avoid exponential growth because these 6 modes are periodic vectors defined by

$$\bar{\mathbf{e}}_i = \exp\left(-\frac{t}{T}\log\lambda_i\right)\cdot\mathbf{e}_i(t)$$

where T is the period of the halo orbit.

For practical implementation it is useful to compute the so-called projection factor Π along $\bar{\mathbf{e}}_1(t)$ which is normal to the hyperplane Γ spanned by the vectors $\bar{\mathbf{e}}_2(t)$ to $\bar{\mathbf{e}}_6(t)$ such that $\bar{\mathbf{e}}_1(t)\cdot\Pi = 1$.

To eliminate the component along the unstable direction $\alpha_1\bar{\mathbf{e}}_1(t)$, we need to add a maneuver $\Delta\mathbf{v}$ so that

$$(\delta\mathcal{E} + \Delta\mathbf{v})\cdot\Pi = 0.$$

That is, the resulting vector $(\delta\mathcal{E} + \Delta\mathbf{v})$ lies in hyperplane Γ and does not have an unstable component.

Hence, $\Delta\mathbf{v}$ needs to be chosen such that

$$\Delta v_x\pi_4 + \Delta v_y\pi_5 + \Delta v_z\pi_6 + \alpha_1 = 0,$$

where π_4, π_5, π_6 are the last 3 components of Π. Choosing a two axis contoller with $\Delta v_z = 0$, and minimizing the Euclidean norm of Δv. We obtain

$$\Delta v_x = -\frac{\alpha_1\pi_4}{\pi_4^2+\pi_5^2}, \qquad \Delta v_y = -\frac{\alpha_1\pi_5}{\pi_4^2+\pi_5^2}.$$

In a similar way, a one or three axis controller can be formulated.

Constaints for Maneuver Implementation. Similar to the Target Point approach, several constraints that impact the maneuvers are specified in the procedure. Some of the most relevant are the time interval between two consecutive tracking epochs (tracking interval), the minimum time interval between maneuvers, and the minimum value of α_1 at which no maneuver will be considered.

- Special emphasis must be placed on the evolution of α_1, examples of which are shown in Figure 8.7.1. With no tracking errors, this parameter increases exponentially with time. When adding tracking errors,

a minimum value $\alpha_{\min}$ which is related to the orbit determination accuracy must be selected in order to prevent a useless maneuver. To avoid a (exponentially increasing) costly maneuver, $\alpha_{\max}$ is chosen: if $\alpha_1 \geq \alpha_{\max}$, a maneuver will be executed. When $\alpha_{\max} > \alpha_1 > \alpha_{\min}$, a maneuver is executed only if an error has been growing exponentially in the previous time step.

- The time between successive maneuvers must be greater than some minimum, $\Delta t_{\min}$. This interval is regulated by orbit determination process, scientific payload requirements and mission operations. Time intervals of 1 to 2 days in the Earth-Moon system are typical.

- If $\Delta \mathbf{v} < \Delta \mathbf{v}_{\min}$ (say, 2 cm/s), the recommended maneuver is cancelled. This is done in order to avoid "small" maneuvers that are of the same order of magnitude as the maneuver implementation errors. It also serves to model actual hardware limitations.

For details of numerical explorations of both Floquet Mode and Target Point methods, see Gómez, Howell, Masdemont, and Simó [1998a].

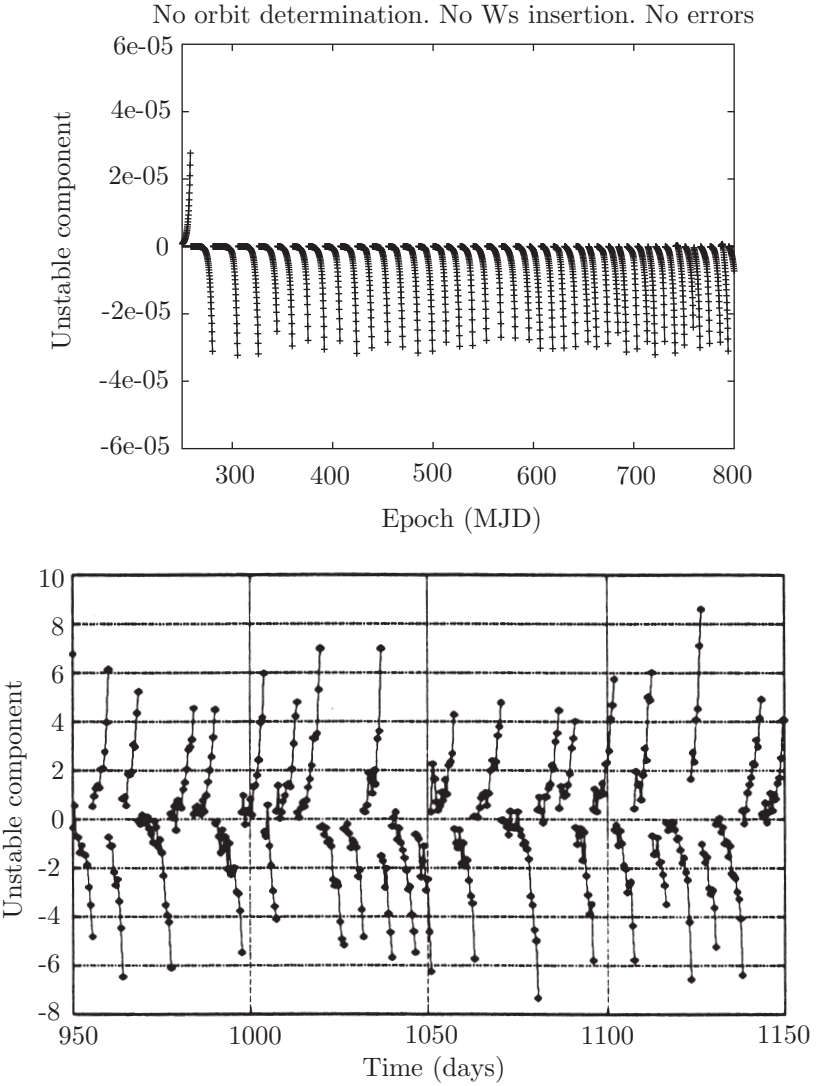


FIGURE 8.7.1. Evolution with time of the unstable component of the controlled orbit. The dotted points correspond to the epochs at which the tracking has been performed. In the upper figure tracking and the maneuvers are performed without error. In the lower figure, errors in both parameters have been introduced (errors still measured in units of 10^{-5}). It is clearly seen, in this last situation, that the unstable component is not so nicely cancelled as in the first one.

9

Invariant Manifolds and Complex Mission Designs

9.1 Introduction

The invariant manifold structures of the collinear libration points for the restricted three-body problem provide the framework for understanding transport phenomena from a geometric point of view. In particular, the stable and unstable invariant manifold tubes associated to libration point orbits are the phase space conduits transporting material between primary bodies for separate three-body systems. Tubes can be used to construct new spacecraft trajectories, such as a "Petit Grand Tour" of the moons of Jupiter. Previous chapters focused on the planar circular restricted three-body problem. The current chapter extends the results to the three-dimensional case.

Besides stable and unstable manifold tubes, center manifolds of the collinear libration points have played a very important role in space mission design. In Chapter 6, we provided a full description of different kinds of libration motions in a large vicinity of these points. In this chapter, we take the next obvious step, and as in Chapter 3, show the existence of heteroclinic connections between pairs of libration orbits, one around the libration point L_1 and the other around L_2. Since these connections are asymptotic orbits, *no maneuver is needed* to perform the transfer from one libration orbit to the other.

Extending Results from Planar Model to Spatial Model. Previous work discussed in earlier chapters, based on the planar circular restricted three-body problem (PCR3BP), revealed the basic structures controlling the dynamics. But actual missions require three-dimensional capabilities, such as control of the latitude and longitude of a spacecraft's escape from and entry into a planetary or moon orbit. For example, a future mission to send a probe to orbit Europa may desire a capture into a high inclination polar orbit around Europa. Three-dimensional capability is also required when decomposing an multibody system into three-body subsystems which are not coplanar, such as the Earth-Sun-spacecraft and Earth-Moon-spacecraft systems (The tilt in the orbital planes of the Earth around the Sun and the Moon around the Earth is about 5 degrees.) These demands necessitate dropping the restriction to planar motion, and extension of earlier results to the spatial model (CR3BP).

In this chapter, we discuss work shown presented in Gómez, Koon, Lo, Marsden, Masdemont, and Ross [2001, 2004]). We show that the invariant manifold structures of the collinear libration points still act as the separatrices between two types of motion: (i) inside the invariant manifold tubes, the motion consists of transit through a neck, the *transit orbits*; (ii) outside the tubes, no such transit motion is possible. We design an algorithm for constructing orbits with any prescribed itinerary and obtain some results for a basic itinerary. Furthermore, we apply these new techniques to the construction of a three-dimensional Petit Grand Tour of the Jovian moon system. By approximating the dynamics of the Jupiter-Europa-Ganymede-spacecraft four-body system as two 3-body subsystems, we seek intersections (in position space) between the tubes of transit orbits enclosed by the stable and unstable manifold tubes.

If the tubes intersect in phase space, we have found a free transfer. Otherwise a ΔV will be necessary to "jump" between the tubes in the two subsystems. As shown in Figure 1.2.9 in Chapter 1, we design an example low energy transfer trajectory from an initial Jovian insertion trajectory, leading to Ganymede and finally to Europa, ending in a high inclination orbit around Europa. It requires a ΔV to jump between the Jupiter-Ganymede and Jupiter-Europa subsystems where the maneuver is labeled in Figure 1.2.9(a).

Computation of the Center Manifold and Its Stable and Unstable Manifolds. In Chapter 6, we established that the linear behavior of collinear libration points is of the type saddle×center×center. This behavior is inherited by the libration orbits, all of which are highly unstable (Gómez, Jorba, Masdemont, and Simó [1991a, 1998b]; Szebehely [1967]). Hence, numerical exploration in the neighborhood of the libration points is not straightforward for two reasons. The first one is high dimensionality of the problem (six phase space dimensions), which makes the explorations rather long, even using simple models like the CR3BP. The second reason

is the highly unstable character of the solutions near the libration points. Due to this instability, errors in the initial conditions multiply by a large factor (on the order of the unstable eigenvector for a halo orbit, $\gg 1$) every half revolution of the secondary around the primary, making it difficult to get an idea of the flow and the orbits in the vicinity of these points.

However, the instability can be handled and the dimensionality reduced by a procedure called *reduction to the center manifold*, to be introduced shortly (see also Gómez, Jorba, Masdemont, and Simó [1991a]; Jorba and Masdemont [1999]). The fundamental idea is based on canonical transformations of the Hamiltonian equations by the Lie series method, implemented in a different way from the "standard" procedure introduced in Deprit [1969]. The change of variables allows us to have a two degree-of-freedom Hamiltonian containing only the orbits in the center manifold. Roughly speaking, this means we remove the main instability. For each level of energy, the orbits are in a three-dimensional manifold that can be represented and viewed globally in a two-dimensional Poincaré surface of section.

The procedure gives a complete description of the libration orbits around an extended neighborhood of the collinear libration points. The main objects found are: planar and vertical families of Lyapunov periodic orbits; Lissajous orbits; periodic halo orbits; and quasi-halo orbits. Apart from using purely numerical procedures, these orbits have also been computed, starting from the equations of motion of the CR3BP and implementing semi-analytic procedures based in asymptotic series of the Lindsted-Poincaré type, as discussed in Chapter 6 and elsewhere (see Gómez, Jorba, Masdemont, and Simó [1991a]; Gómez, Masdemont, and Simó [1997, 1998]; Jorba and Masdemont [1999]).

In this chapter we implement a slightly different procedure for the computation of the reduction to the center manifold, in order to have all the possible initial conditions in the center manifold of a selected level of energy. Then, the connections between the orbits of L_1 and L_2 are constructed looking for the intersections of the unstable manifold of a libration orbit around L_i with the stable one of a libration one around L_{3-i} for $i = 1, 2$. The orbits are asymptotic to both libration orbits in the same level of energy and thus, in the ideal situation of the CR3BP, no ΔV is required to perform the transfer from one orbit to the other.

This efficient way of computing stable and unstable manifolds for the center manifold of a selected level of energy allows us to construct the Petit Grand Tour of Jovian moons.

9.2 The Linearized Hamiltonian System

Studying the linearization of the dynamics near the equilibria is of course an essential ingredient for understanding the more complete nonlinear dynam-

ics (Conley [1968]; McGehee [1969]; Appleyard [1970]; Koon, Lo, Marsden, and Ross [2000]). In fact, it can be shown that for a value of Jacobi constant just below that of L_1 (L_2), the nonlinear dynamics in the equilibrium region $\mathcal{R}_1$ ($\mathcal{R}_2$) is qualitatively the same as the linearized picture that we will describe below. For details, see §9.7 as well as other references (Gómez and Masdemont [2000]; Jorba and Masdemont [1999]; Hartman [1964]; Wiggins, Wiesenfeld, Jaffé, and Uzer [2001]). This geometric insight will be used later to guide our numerical explorations in constructing orbits with prescribed itineraries.

Expansion of CR3BP Hamiltonian around Collinear Equilibria. From the work of Chapter 6, the Hamiltonian for the restricted three-body problem has the form

$$H = \frac{1}{2}\left(p_x^2 + p_y^2 + p_z^2\right) + yp_x - xp_y - \frac{1-\mu}{r_1} - \frac{\mu}{r_2}.$$

Moreover, its expansion around a collinear equilibrium point is given by

$$H = \frac{1}{2}\left(p_x^2 + p_y^2 + p_z^2\right) + yp_x - xp_y - \sum_{n\geq 2} c_n(\mu)\rho^n P_n\left(\frac{x}{\rho}\right). \qquad (9.2.1)$$

where $\rho^2 = x^2 + y^2 + z^2$ and P_n is the Legendre polynomial of degree n. The coefficients c_n are given by

$$c_n = \frac{1}{\gamma_j^3}\left((\pm 1)^n \mu + (-1)^n \frac{(1-\mu)\gamma_j^{n+1}}{(1\mp\gamma_j)^{n+1}}\right), \quad \text{for } L_j,\ j = 1, 2$$

where γ_j is the distance between L_j and the second primary. As usual, the upper sign is for L_1 and the lower one for L_2.

This expression for the Hamiltonian is in an ideal form because all orders of the nonlinear expansion can be developed recursively using the well-known Legendre polynomial relationships. For instance, if we define

$$T_n(x, y, z) = \rho^n P_n\left(\frac{x}{\rho}\right),$$

then T_n is a homogeneous polynomial of degree n that satisfies the following recurrence relations

$$T_n = \frac{2n-1}{n} x T_{n-1} - \frac{n-1}{n}(x^2 + y^2 + z^2) T_{n-2},$$

which start with $T_0 = 1$ and $T_1 = x$. This is particularly useful if the successive approximation procedure is carried to high orders via algebraic manipulation software.

Normal Form for the Linearized Hamiltonian. The linearization around the equilibrium point is given by the second-order terms of the CR3BP Hamiltonian

$$H_2 = \frac{1}{2}\left(p_x^2 + p_y^2\right) + yp_x - xp_y - c_2x^2 + \frac{c_2}{2}y^2 + \frac{1}{2}p_z^2 + \frac{c_2}{2}z^2. \tag{9.2.2}$$

It will be shown below that the local behavior near these points are of the type saddle×center×center, and a real linear and symplectic change of coordinates can be found to cast the second-order part of the Hamiltonian (9.2.2) into its real normal form,

$$H_2 = \lambda q_1 p_1 + \frac{\omega_p}{2}(q_2^2 + p_2^2) + \frac{\omega_v}{2}(q_3^2 + p_3^2). \tag{9.2.3}$$

This normal form will be used in performing qualitative analysis in §9.3 and can also be considered as a first step of normal form computations for the full Hamiltonian. Here, λ, ω_p and ω_v are positive real numbers given by

$$\lambda^2 = \frac{c_2 - 2 + \sqrt{9c_2^2 - 8c_2}}{2}, \qquad \omega_p^2 = \frac{2 - c_2 + \sqrt{9c_2^2 - 8c_2}}{2}, \qquad \omega_v^2 = c_2.$$

Since $c_2 > 1$ for collinear points, the vertical direction in (9.2.2) is an harmonic oscillator with frequency $\omega_v = \sqrt{c_2}$. As the vertical direction is already uncoupled from the planar ones, in what follows we will focus first on the planar directions, i.e.,

$$H_2 = \frac{1}{2}\left(p_x^2 + p_y^2\right) + yp_x - xp_y - c_2x^2 + \frac{c_2}{2}y^2. \tag{9.2.4}$$

For simplicity, we keep the same name H_2 for the Hamiltonian. The next step is to compute a *symplectic* change of variables such that the Hamiltonian (9.2.4) is in normal form.

Symplectic Matrix and Canonical Transformation. The canonical ***Hamiltonian equations*** for an autonomous system with ***Hamiltonian*** $H(q,p) = H(z)$, where $z = (q,p) = (q_1, \ldots, q_n; p_1, \ldots, p_n) \in \mathbb{R}^{2n}$ can be written in a compact form as follows:

$$\dot{z} = J\nabla H(z). \tag{9.2.5}$$

Here, $\nabla = (\partial_{z_1}, \ldots, \partial_{z_{2n}})$ is the gradient operator on $\mathbb{R}^{2n}$, and J is the $2n \times 2n$ symplectic matrix

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

where I_n denotes the $n \times n$ identity matrix.

Since J satisfies $J^T = -J$, it defines a symplectic form Ω on $\mathbb{R}^{2n}$:

$$\Omega(u, v) = u^T J v, \qquad u, v \in R^{2n},$$

which is a non-degenerate bilinear skew symmetric form. A matrix M is said to be *symplectic* if it satisfies

$$M^T J M = J. \tag{9.2.6}$$

Notice that J is symplectic since $J^{-1} = J^T$. The right hand side of (9.2.5) is called the symplectic gradient of H.

Now let us consider the effect that a change of variables,

$$\bar{z} = \phi(z), \tag{9.2.7}$$

has on a Hamiltonian system associated with H. Note that the Hamiltonian equations obtained from the Hamiltonian $H \circ \phi$ can be different from the equations obtained by applying the transformation ϕ to the Hamiltonian equations associated with H. When these two systems of differential equations coincide, it is said that the transformation ϕ preserves the Hamiltonian character of the system.

A change of variables is called *canonical* when it preserves the Hamiltonian character of the system. Canonical transformations are very useful both from the theoretical and the practical points of view, since they allow one to work on a single function, the Hamiltonian, instead of a system of $2n$ differential equations. It can be shown as shown that a transformation is canonical if and only the differential of the change at any point is a symplectic matrix.

Differentiate equation (9.2.7), we obtain

$$\dot{\bar{z}} = \frac{\partial \phi}{\partial z} \dot{z}. \tag{9.2.8}$$

Substituting equation (9.2.5) into (9.2.8) and apply the chain rule, we have

$$\begin{aligned}
\dot{\bar{z}} &= \frac{\partial \phi}{\partial z} \dot{z} \\
&= \frac{\partial \phi}{\partial z} J \nabla H(z) \\
&= \frac{\partial \phi}{\partial z} J \left(\frac{\partial H}{\partial z}(z) \right)^T \\
&= \frac{\partial \phi}{\partial z} J \left(\frac{\partial H}{\partial \bar{z}}(\bar{z}) \frac{\partial \phi}{\partial z} \right)^T \\
&= \frac{\partial \phi}{\partial z} J \left(\frac{\partial \phi}{\partial z} \right)^T \left(\frac{\partial H}{\partial \bar{z}}(\bar{z}) \right)^T
\end{aligned}$$

$$= \frac{\partial \phi}{\partial z} J \left(\frac{\partial \phi}{\partial z} \right)^T \nabla H(\bar{z})$$

Hence, the coordinate change, $\bar{z} = \phi(z)$, preserves the Hamiltonian character if and only if

$$\frac{\partial \phi}{\partial z} J \left(\frac{\partial \phi}{\partial z} \right)^T = J. \tag{9.2.9}$$

This proves our claim since it is easy to show that

$$MJM^T = J \tag{9.2.10}$$

is equivalent to equation (9.2.6) by taking the inverse of (9.2.10) and by knowing the fact that symplectic matrices form a group.

For a linearized Hamiltonian system associated with H_2 (see (9.2.4)), a symplectic linear coordinate change is clearly canonical.

Eigenvalues. The equations of motion associated the Hamiltonian (9.2.4) are given by the linear system

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{p_x} \\ \dot{p_y} \end{pmatrix} = J \nabla H_2 = J \text{Hess}(H_2) \begin{pmatrix} x \\ y \\ p_x \\ p_y \end{pmatrix}.$$

Define $J\text{Hess}(H_2)$ as matrix M

$$M = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 2c_2 & 0 & 0 & 1 \\ 0 & -c_2 & -1 & 0 \end{pmatrix}. \tag{9.2.11}$$

Similar to §2.7, the characteristic polynomial is

$$p(\beta) = \beta^4 + (2 - c_2)\beta^2 + (1 + c_2 - 2c_2^2). \tag{9.2.12}$$

Let $\alpha = \beta^2$, then the roots of $p(\alpha) = 0$ are as follows

$$\alpha_1 = \frac{c_2 - 2 + \sqrt{9c_2^2 - 8c_2}}{2}, \qquad \alpha_2 = \frac{c_2 - 2 - \sqrt{9c_2^2 - 8c_2}}{2}.$$

Since $c_2 > 1$ and the last term of $p(\alpha) = 0$ is negative, this quadratic equation must have one positive and one negative root. So, we have $\alpha_1 > 0$ and $\alpha_2 < 0$. This shows that the equilibrium point is a saddle×center×center. Thus, let us define λ as $\sqrt{\alpha_1}$ and ω_p as $\sqrt{-\alpha_2}$.

Eigenvectors. Now, we want to find the eigenvectors of matrix (9.2.11) and use them to construct a symplectic linear change of variables which will cast (9.2.4) into its real normal form. In doing so, we will take advantage of the special form of this matrix. Denote the matrix $M - \beta I_4$ by M_β, then

$$M_\beta = \begin{pmatrix} A_\beta & I_2 \\ B & A_\beta \end{pmatrix}, \quad A_\beta = \begin{pmatrix} -\beta & 1 \\ -1 & -\beta \end{pmatrix}, \quad B = \begin{pmatrix} 2c_2 & 0 \\ 0 & -c_2 \end{pmatrix}.$$

Now, the elements of the kernel, denoted as (k_1, k_2), can be found by first solving

$$(B - A_\beta^2)k_1 = 0,$$

and then

$$k_2 = -A_\beta k_1.$$

Thus, the (column) eigenvectors of M are given by

$$(2\beta, \beta^2 - 2c_2 - 1, \beta^2 + 2c_2 + 1, \beta^3 + (1 - 2c_2)\beta), \tag{9.2.13}$$

where β denotes one of the eigenvalues.

Let us start with the eigenvectors related to $i\omega_p$. From (9.2.12), we obtain

$$\omega_p^4 - (2 - c_2)\omega_p^2 + (1 + c_2 - 2c_2^2) = 0.$$

After substituting $\beta = i\omega_p$ to the expression of the eigenvector (9.2.13) and separating real and imaginary parts as $u_{\omega_p} + iv_{\omega_p}$, we obtain two eigenvectors

$$\begin{aligned} u_{\omega_p} &= (0, -\omega_p^2 - 2c_2 - 1, -\omega_p^2 + 2c_2 + 1, 0), \\ v_{\omega_p} &= (2\omega_p, 0, 0, -\omega_p^3 + (1 - 2c_2)\omega_p). \end{aligned}$$

Moreover, the rest of eigenvectors associated with eigenvalues $\pm\lambda$ can also be obtained similarly

$$\begin{aligned} u_{+\lambda} &= (2\lambda, \lambda^2 - 2c_2 - 1, \lambda^2 + 2c_2 + 1, \lambda^3 + (1 - 2c_2)\lambda), \\ u_{-\lambda} &= (-2\lambda, \lambda^2 - 2c_2 - 1, \lambda^2 + 2c_2 + 1, -\lambda^3 - (1 - 2c_2)\lambda). \end{aligned}$$

Symplectic Change of Variables. Initially, we consider the change of variables defined by the matrix C

$$C = (u_{+\lambda}, u_{\omega_p}, u_{-\lambda}, v_{\omega_p})$$

which comprises the four eigenvectors. To find out whether this matrix is symplectic or not, we check $C^T J C = J$. It is tedious but not difficult to see that

$$C^T J C = \begin{pmatrix} 0 & D \\ -D & 0 \end{pmatrix}, \quad D = \begin{pmatrix} d_\lambda & 0 \\ 0 & d_{\omega_p} \end{pmatrix},$$

where

$$d_\lambda = 2\lambda((4+3c_2)\lambda^2 + 4 + 5c_2 - 6c_2^2),$$
$$d_{\omega_p} = \omega_p((4+3c_2)\omega_p^2 - 4 - 5c_2 + 6c_2^2).$$

This implies that we need to apply some scaling on the columns of C in order to have a symplectic change. Since it can be shown that $d_\lambda > 0$ and $d_{\omega_p} > 0$ for $0 < \mu \leq \frac{1}{2}$, the scaling is given by the factors $s_1 = \sqrt{d_\lambda}$ and $s_2 = \sqrt{d_{\omega_p}}$.

To obtain the final change, we have to take into account the vertical direction (z, p_z): to put it into real normal form we use the substitution

$$z = \frac{1}{\sqrt{\omega_v}} q_3, \qquad p_z = \sqrt{\omega_v} p_3.$$

This implies that the final change is given by the symplectic matrix C

$$= \begin{pmatrix} \frac{2\lambda}{s_1} & 0 & 0 & \frac{-2\lambda}{s_1} & \frac{2\omega_p}{s_2} & 0 \\ \frac{\lambda^2-2c_2-1}{s_1} & \frac{-\omega_p^2-2c_2-1}{s_2} & 0 & \frac{\lambda^2-2c_2-1}{s_1} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{\omega_v}} & 0 & 0 & 0 \\ \frac{\lambda^2+2c_2+1}{s_1} & \frac{-\omega_p^2+2c_2+1}{s_2} & 0 & \frac{\lambda^2+2c_2+1}{s_1} & 0 & 0 \\ \frac{\lambda^3+(1-2c_2)\lambda}{s_1} & 0 & 0 & \frac{-\lambda^3-(1-2c_2)\lambda}{s_1} & \frac{-\omega_p^3+(1-2c_2)\omega_p}{s_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{\omega_v} \end{pmatrix}, \tag{9.2.14}$$

that casts Hamiltonian (9.2.2) into its real normal form,

$$H_2 = \lambda q_1 p_1 + \frac{\omega_p}{2}(q_2^2 + p_2^2) + \frac{\omega_v}{2}(q_3^2 + p_3^2). \tag{9.2.15}$$

A short computation gives the linearized equations in the form

$$\begin{aligned} \dot{q}_1 &= \lambda q_1, & \dot{p}_1 &= -\lambda p_1, \\ \dot{q}_2 &= \omega_p p_2, & \dot{p}_2 &= -\omega_p q_2, \\ \dot{q}_3 &= \omega_v p_3, & \dot{p}_3 &= -\omega_v q_3. \end{aligned} \tag{9.2.16}$$

Solutions of the equations (9.2.16) can be conveniently written as

$$\begin{aligned} q_1(t) &= q_1^0 e^{\lambda t}, \qquad p_1(t) = p_1^0 e^{-\lambda t}, \\ q_2(t) + ip_2(t) &= (q_2^0 + ip_2^0) e^{-i\omega_p t}, \\ q_3(t) + ip_2(t) &= (q_3^0 + ip_3^0) e^{-i\omega_v t}, \end{aligned} \tag{9.2.17}$$

where the constants q_1^0, p_1^0, $q_2^0 + ip_2^0$, and $q_3^0 + ip_3^0$ are the initial conditions. These linearized equations admit integrals in addition to the Hamiltonian function; namely, the functions $q_1 p_1$, $q_2^2 + p_2^2$ and $q_3^2 + p_3^2$ are constant along solutions.

9.3 Invariant Manifold as Separatrix

The Linearized Phase Space. For positive h and c, the region $\mathcal{R}$, which is determined by

$$H_2 = h, \quad \text{and} \quad |p_1 - q_1| \leq c,$$

is homeomorphic to the product of a 4-sphere and an interval I, $S^4 \times I$; namely, for each fixed value of $p_1 - q_1$ in the interval $[-c, c]$, we see that the equation $H_2 = h$ determines a 4-sphere

$$\frac{\lambda}{4}(q_1 + p_1)^2 + \frac{\nu}{2}(q_2^2 + p_2^2) + \frac{\omega}{2}(q_3^2 + p_3^2) = h + \frac{\lambda}{4}(p_1 - q_1)^2.$$

The bounding 4-sphere of $\mathcal{R}$ for which $p_1 - q_1 = -c$ will be called n_1, and that where $p_1 - q_1 = c$, n_2 (see Figure 9.3.1).

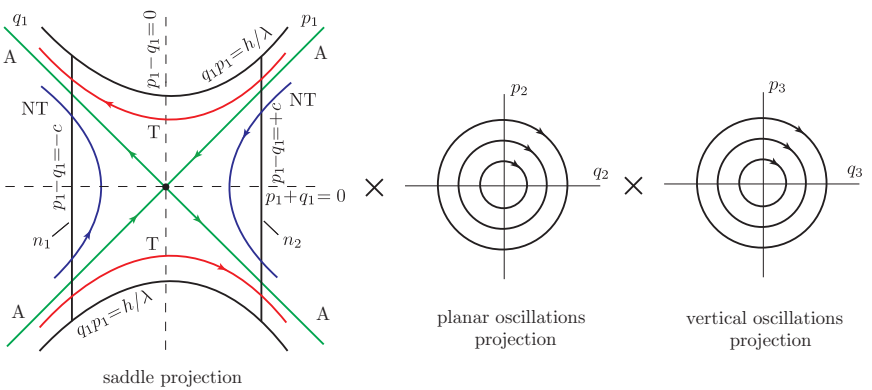


FIGURE 9.3.1. The flow in the equilibrium region has the form saddle × center × center. On the left is shown the projection onto the (p_1, q_1)-plane (note, axes tilted 45°). Shown are the bounded orbits (black dot at the center), the asymptotic orbits (labeled A), two transit orbits (T) and two non-transit orbits (NT).

Similar to Chapter 2, we call the set of points on each bounding 4-sphere where $q_1 + p_1 = 0$ the ***equator***, and the sets where $q_1 + p_1 > 0$ or $q_1 + p_1 < 0$ will be called the ***north*** and ***south hemispheres***, respectively.

The Linear Flow in $\mathcal{R}$. To analyze the flow in $\mathcal{R}$, one considers the projections on the (q_1, p_1)-plane and $(q_2, p_2) \times (q_3, p_3)$-space, respectively. In the first case we see the standard picture of an unstable critical point, and in the second, of a center consisting of two uncoupled harmonic oscillators. Figure 9.3.1 schematically illustrates the flow. The coordinate axes of the (q_1, p_1)-plane have been tilted by 45° and labeled (p_1, q_1) in order

to correspond to the direction of the flow in later figures which adopt the NASA convention that the larger primary is to the left of the smaller secondary. With regard to the first projection we see that $\mathcal{R}$ itself projects to a set bounded on two sides by the hyperbola $q_1 p_1 = h/\lambda$ (corresponding to $q_2^2 + p_2^2 = q_3^2 + p_3^2 = 0$, see (2.7.4)) and on two other sides by the line segments $p_1 - q_1 = \pm c$, which correspond to the bounding 4-spheres.

Since $q_1 p_1$ is an integral of the equations in $\mathcal{R}$, the projections of orbits in the (q_1, p_1)-plane move on the branches of the corresponding hyperbolas $q_1 p_1 =$ constant, except in the case $q_1 p_1 = 0$, where $q_1 = 0$ or $p_1 = 0$. If $q_1 p_1 > 0$, the branches connect the bounding line segments $p_1 - q_1 = \pm c$ and if $q_1 p_1 < 0$, they have both end points on the same segment. A check of equation (9.2.17) shows that the orbits move as indicated by the arrows in Figure 9.3.1.

To interpret Figure 9.3.1 as a flow in $\mathcal{R}$, notice that each point in the (q_1, p_1)-plane projection corresponds to a 3-sphere S^3 in $\mathcal{R}$ given by

$$\frac{\nu}{2}(q_2^2 + p_2^2) + \frac{\omega}{2}(q_3^2 + p_3^2) = h - \lambda q_1 p_1.$$

Of course, for points on the bounding hyperbolic segments ($q_1 p_1 = h/\lambda$), the 3-sphere collapses to a point. Thus, the segments of the lines $p_1 - q_1 = \pm c$ in the projection correspond to the 4-spheres bounding $\mathcal{R}$. This is because each corresponds to a 3-sphere crossed with an interval where the two end 3-spheres are pinched to a point.

We distinguish nine classes of orbits grouped into the following four categories:

1. The point $q_1 = p_1 = 0$ corresponds to an invariant 3-sphere S_h^3 of ***bounded orbits*** (periodic and quasi-periodic) in $\mathcal{R}$. This 3-sphere is given by

$$\frac{\nu}{2}(q_2^2 + p_2^2) + \frac{\omega}{2}(q_3^2 + p_3^2) = h, \qquad q_1 = p_1 = 0. \tag{9.3.1}$$

   It is an example of a ***normally hyperbolic invariant manifold*** (NHIM) (see Wiggins [1994]). Roughly, this means that the stretching and contraction rates under the linearized dynamics transverse to the 3-sphere dominate those tangent to the 3-sphere. This is clear for this example since the dynamics normal to the 3-sphere are described by the exponential contraction and expansion of the saddle point dynamics. Here the 3-sphere acts as a "big saddle point". See the black dot at the center of the (q_1, p_1)-plane on the left side of Figure 9.3.1.

2. The four half open segments on the axes, $q_1 p_1 = 0$, correspond to four cylinders of orbits asymptotic to this invariant 3-sphere S_h^3 either as time increases ($p_1 = 0$) or as time decreases ($q_1 = 0$). These are

called ***asymptotic*** orbits and they form the stable and the unstable manifolds of S_h^3. The ***stable manifolds***, $W_{\pm}^s(S_h^3)$, are given by

$$\frac{\nu}{2}(q_2^2+p_2^2)+\frac{\omega}{2}(q_3^2+p_3^2)=h, \qquad q_1=0. \tag{9.3.2}$$

$W_+^s(S_h^3)$ (with $p_1>0$) is the branch going from right to left and $W_-^s(S_h^3)$ (with $p_1<0$) is the branch going from left to right. The unstable manifolds, $W_{\pm}^u(S_h^3)$, are given by

$$\frac{\nu}{2}(q_2^2+p_2^2)+\frac{\omega}{2}(q_3^2+p_3^2)=h, \qquad p_1=0. \tag{9.3.3}$$

$W_+^u(S_h^3)$ (with $q_1>0$) is the branch going from right to left and $W_-^u(S_h^3)$ (with $q_1<0$) is the branch going from left to right. See the four orbits labeled A of Figure 9.3.1.

3. The hyperbolic segments determined by $q_1p_1=$ constant >0 correspond to two cylinders of orbits which cross $\mathcal{R}$ from one bounding 4-sphere to the other, meeting both in the same hemisphere; the northern hemisphere if they go from $p_1-q_1=+c$ to $p_1-q_1=-c$, and the southern hemisphere in the other case. Since these orbits transit from one realm to another, we call them ***transit*** orbits. See the two orbits labeled T of Figure 9.3.1.

4. Finally the hyperbolic segments determined by $q_1p_1=$ constant <0 correspond to two cylinders of orbits in $\mathcal{R}$ each of which runs from one hemisphere to the other hemisphere on the same bounding 4-sphere. Thus if $q_1>0$, the 4-sphere is n_1 ($p_1-q_1=-c$) and orbits run from the southern hemisphere ($q_1+p_1<0$) to the northern hemisphere ($q_1+p_1>0$) while the converse holds if $q_1<0$, where the 4-sphere is n_2. Since these orbits return to the same realm, we call them ***non-transit*** orbits. See the two orbits labeled NT of Figure 9.3.1.

McGehee Representation. As noted above, $\mathcal{R}$ is a 5-dimensional manifold that is homeomorphic to $S^4\times I$. It can be represented by a spherical annulus bounded by two 4-spheres n_1, n_2, as shown in Figure 9.3.2(b).

Figure 9.3.2(a) is a cross-section of $\mathcal{R}$. Notice that this cross-section is qualitatively the same as the illustration in Figure 9.3.1. The following classifications of orbits correspond to the previous four categories:

1. There is an invariant 3-sphere S_h^3 of bounded orbits in the region $\mathcal{R}$ corresponding to the black dot in the middle of Figure 9.3.2(a). Notice that this 3-sphere is the equator of the central 4-sphere given by $p_1-q_1=0$.

2. Again let n_1, n_2 be the bounding 4-spheres of region $\mathcal{R}$, and let n denote either n_1 or n_2. We can divide n into two hemispheres: n^+,

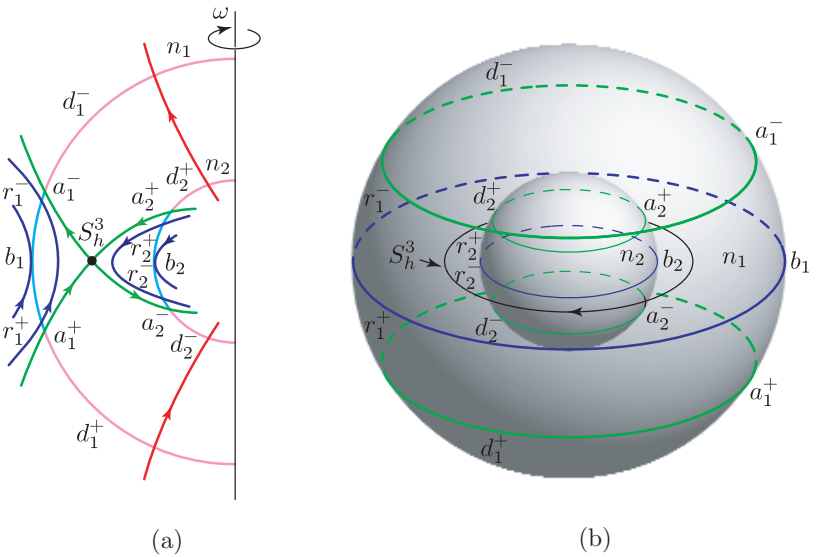


FIGURE 9.3.2. (a) The cross-section of the flow in the $\mathcal{R}$ region of the energy surface. (b) The McGehee representation of the flow in the region $\mathcal{R}$.

where the flow enters $\mathcal{R}$, and n^-, where the flow leaves $\mathcal{R}$. There are four cylinders of orbits *asymptotic* to the invariant 3-sphere S_h^3. They form the stable and unstable manifolds to the invariant 3-sphere S_h^3. Topologically, both invariant manifolds look like 4-dimensional "tubes" ($S^3 \times \mathbb{R}$) inside a 5-dimensional energy manifold. The interior of the stable manifolds $W_{\pm}^s(S_h^3)$ and unstable manifolds $W_{\pm}^s(S_h^3)$ can be given as follows

$$\begin{aligned}
\text{int}(W_+^s(S_h^3)) &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad p_1 > q_1 > 0\},\\
\text{int}(W_-^s(S_h^3)) &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad p_1 < q_1 < 0\},\\
\text{int}(W_+^u(S_h^3)) &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad q_1 > p_1 > 0\},\\
\text{int}(W_-^u(S_h^3)) &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad q_1 < p_1 < 0\}.
\end{aligned} \tag{9.3.4}$$

The exterior of these invariant manifolds can be given similarly from studying Figure 9.3.2(a).

3. Let a^+ and a^- (where $q_1 = 0$ and $p_1 = 0$ respectively) be the intersections of the stable and unstable manifolds with the bounding sphere n. Then a^+ appears as a 3-sphere in n^+, and a^- appears as a 3-sphere in n^-. Consider the two spherical caps on each bounding 4-sphere given by

$$\begin{aligned}
d_1^+ &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad p_1 - q_1 = -c, \quad p_1 < q_1 < 0\},\\
d_1^- &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad p_1 - q_1 = -c, \quad q_1 > p_1 > 0\},\\
d_2^+ &= \{(q_1,p_1,q_2,p_2,q_3,p_3) \in \mathcal{R} \mid \quad p_1 - q_1 = +c, \quad p_1 > q_1 > 0\},
\end{aligned}$$

$$d_2^- = \{(q_1, p_1, q_2, p_2, q_3, p_3) \in \mathcal{R} \mid \quad p_1 - q_1 = +c, \quad q_1 < p_1 < 0\}.$$

Since d_1^+ is the spherical cap in n_1^+ bounded by a_1^+, then the *transit* orbits entering $\mathcal{R}$ on d_1^+ exit on d_2^- of the other bounding sphere. Similarly, since d_1^- is the spherical cap in n_1^- bounded by a_1^-, the transit orbits leaving on d_1^- have come from d_2^+ on the other bounding sphere. Note that all spherical caps where the transit orbits pass through are in the interior of stable and unstable manifold tubes.

4. Let b be the intersection b of n^+ and n^- (where $q_1 + p_1 = 0$). Then, b is a 3-sphere of tangency points. Orbits tangent at this 3-sphere "bounce off," i.e., do not enter $\mathcal{R}$ locally. Moreover, if we let r^+ be a spherical zone which is bounded by a^+ and b, then *non-transit* orbits entering $\mathcal{R}$ on r^+ exit on the same bounding 4-sphere through r^- which is bounded by a^- and b. It is easy to show that all the spherical zones where non-transit orbits bounce off are in the exterior of stable and unstable manifold tubes.

Invariant Manifolds as Separatrices. The key observation here is that the asymptotic orbits form 4-dimensional stable and unstable manifold "tubes" ($S^3 \times \mathbb{R}$) to the invariant 3-sphere S_h^3 in a 5-dimensional energy surface and they separate two distinct types of motion: transit orbits and non-transit orbits. The transit orbits, passing from one realm to another, are those inside the 4-dimensional manifold tube. The non-transit orbits, which bounce back to their realm of origin, are those outside the tube.

In fact, it can be shown that for a value of Jacobi constant just below that of L_1 (L_2), the nonlinear dynamics in the equilibrium region $\mathcal{R}_1$ ($\mathcal{R}_2$) is qualitatively the same as the linearized picture that we have shown above.

For example, the normally hyperbolic invariant manifold (NHIM) for the nonlinear system which corresponds to the 3-sphere (9.3.1) for the linearized system is given by

$$\mathcal{M}_h^3 = \left\{(q,p) \mid \quad \frac{\nu}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega}{2}\left(q_3^2 + p_3^2\right) + f(q_2, p_2, q_3, p_3) = h, \quad q_1 = p_1 = 0\right\} \tag{9.3.5}$$

where f is at least of third-order. Here, $(q_1, p_1, q_2, p_2, q_3, p_3)$ are normal form coordinates and are related to the linearized coordinates via a near-identity transformation.

In a small neigborhood of the equilibrium point, since the nonlinear terms are much smaller than the linear terms, the 3-sphere for the linear problem becomes a deformed sphere for the nonlinear problem. Moreover, since NHIMs persist under perturbation, this deformed sphere $\mathcal{M}_h^3$ still has stable and unstable manifolds which are given by

$$W_{\pm}^s(\mathcal{M}_h^3) = \left\{(q,p) \mid \quad \frac{\nu}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega}{2}\left(q_3^2 + p_3^2\right) + f(q_2, p_2, q_3, p_3) = h, \quad q_1 = 0\right\}$$

$$W_{\pm}^u(\mathcal{M}_h^3) = \left\{(q,p) \mid \quad \frac{\nu}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega}{2}\left(q_3^2 + p_3^2\right) + f(q_2, p_2, q_3, p_3) = h, \quad p_1 = 0\right\}.$$

Notice the similarity between the formulas above and those for the linearized problem (9.3.2 and 9.3.3),

See §9.6-9.8 at the end of this chapter as well as other references (Gómez and Masdemont [2000]; Jorba and Masdemont [1999]; Hartman [1964]; Wiggins, Wiesenfeld, Jaffé, and Uzer [2001]) for details. This geometric insight will be used below to guide our numerical explorations in constructing orbits with prescribed itineraries.

9.4 Construction of Orbits with Prescribed Itineraries in the Spatial Case

Recall the Construction in the Planar Case. In earlier chapters on the planar case, a numerical demonstration is given of a heteroclinic connection between pairs of equal Jacobi constant Lyapunov orbits, one around L_1, the other around L_2. This heteroclinic connection augments the homoclinic orbits associated with the L_1 and L_2 Lyapunov orbits, which were previously known (McGehee [1969]). Linking these heteroclinic connections and homoclinic orbits leads to *dynamical chains*.

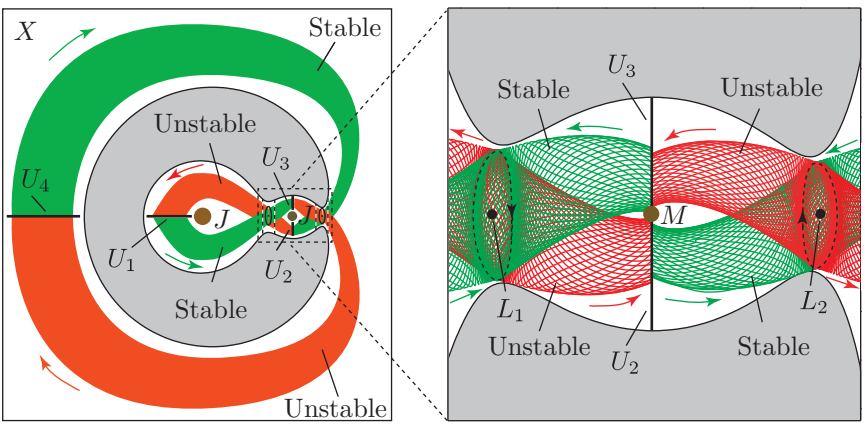


FIGURE 9.4.1. Location of libration point orbit invariant manifold tubes in position space. Stable manifolds are lightly shaded, unstable manifolds are darkly. The location of the Poincaré sections (U_1, U_2, U_3, and U_4) are also shown.

The dynamics in the neighborhood of these chains gives rise to interesting analytical results. We proved the existence of a large class of interesting orbits near a chain which a spacecraft can follow in its rapid transition between the inside and outside of a Jovian moon's orbit via a moon encounter. The global collection of these orbits is called a *dynamical channel*. We proved a theorem which gives the global orbit structure in the neighborhood of a chain. In simplified form, the theorem essentially says:

For any admissible bi-infinite sequence $(\ldots, u_{-1}; u_0, u_1, u_2, \ldots)$ *of symbols* $\{I, M, X\}$ *where* I*,* M*, and* X *stand for the interior, moon, and exterior realms respectively, there corresponds an orbit near the chain whose past and future whereabouts with respect to these three realms match those of the given sequence.*

For example, consider the Jupiter-Ganymede-spacecraft 3-body system. Given the bi-infinite sequence $(\ldots, I; M, X, M, \ldots)$, there exists an orbit starting in the Ganymede realm which came from the interior realm and is going to the exterior realm and returning to the Ganymede realm.

Moreover, we not only proved the existence of orbits with prescribed itineraries, but developed a systematic procedure for their numerical construction. We will illustrate below the numerical construction of orbits with prescribed finite (but large enough) itineraries in the three-body planet-moon-spacecraft problem. As our example, chosen for simplicity of exposition, we construct a spacecraft orbit with the central block $(M, X; M, I, M)$.

Example Itinerary: $(M, X; M, I, M)$**.** For the present numerical construction, we adopt the following convention. The U_1 and U_4 Poincaré sections will be $(y = 0, x < 0, \dot{y} < 0)$ in the interior realm, and $(y = 0, x < -1, \dot{y} > 0)$ in the exterior realm, respectively. The U_2 and U_3 sections will be $(x = 1 - \mu, y < 0, \dot{x} > 0)$ and $(x = 1 - \mu, y > 0, \dot{x} < 0)$ in the moon realm, respectively. See Figure 9.4.1 for the location of the Poincaré sections relative to the tubes.

A key observation for the planar case is a result which has shown that the invariant manifold tubes separate two types of motion. See Figures 9.4.2(a) and 9.4.2(b). The orbits inside the tube transit from one realm to another; those outside the tubes bounce back to their original realm.

Since the upper curve in Figure 9.4.2(b) is the Poincaré cut of the stable manifold of the periodic orbit around L_1 in the U_3 plane, a point inside that curve is an orbit that goes from the moon realm to the interior realm, so this region can be described by the label $(; M, I)$. Similarly, a point inside the lower curve of Figure 9.4.2(b) came from the exterior realm into the moon realm, and so has the label $(X; M)$. A point inside the intersection $\Delta_{\mathcal{M}}$ of *both* curves is an $(X; M, I)$ orbit, so it makes a transition from the exterior realm to the interior realm, passing through the moon realm. Similarly, by choosing Poincaré sections in the interior and the exterior realm, i.e., in the U_1 and U_4 plane, we find the intersection region $\Delta_{\mathcal{I}}$ consisting of $(M; I, M)$ orbits, and $\Delta_{\mathcal{X}}$, which consists of $(M; X, M)$ orbits.

Flowing the intersection $\Delta_{\mathcal{X}}$ forward to the moon realm, it stretches into the strips in Figure 9.4.2(c). These strips are the image of $\Delta_{\mathcal{X}}$ (i.e., $P(\Delta_{\mathcal{X}})$) under the Poincaré map P, and thus get the label $(M, X; M)$. Similarly, flowing the intersection Δ_I backward to the moon realm, it stretches into the strips $P^{-1}(\Delta_{\mathcal{I}})$ in Figure 9.4.2(c), and thus have the label $(; M, I, M)$. The intersection of these two types of strips (i.e., $\Delta_{\mathcal{M}} \cap P(\Delta_{\mathcal{X}}) \cap P^{-1}(\Delta_{\mathcal{I}})$) consist of the desired $(M, X; M, I, M)$ orbits. If we take any point inside

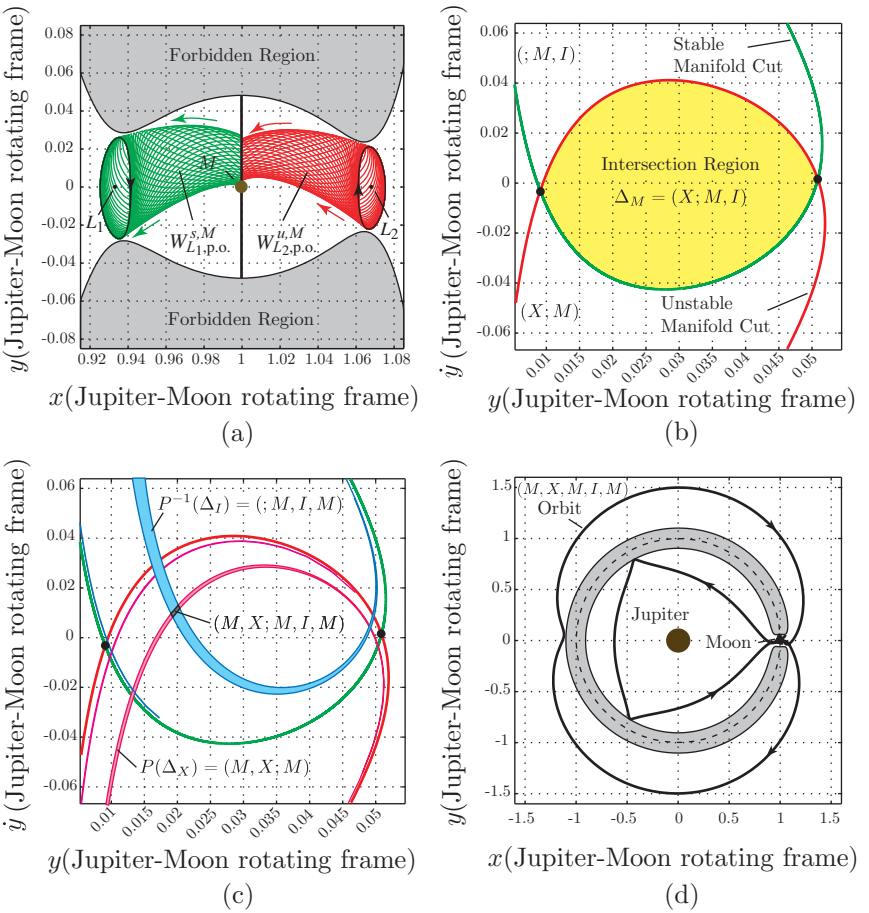


FIGURE 9.4.2. (a) The projection of invariant manifolds $W^{s,\mathcal{M}}_{L_1,\text{p.o.}}$ and $W^{u,\mathcal{M}}_{L_2,\text{p.o.}}$ in the realm M of the position space. (b) A close-up of the intersection realm between the Poincaré cuts of the invariant manifolds on the U_3 section ($x = 1-\mu, y > 0$). (c) Intersection between image of Δ_X and pre-image of Δ_I labeled $(M,X;M,I,M)$. (d) Example orbit passing through $(M,X;M,I,M)$ region of (c).

these intersections and integrate it forward and backward, we find the desired orbits. See Figure 9.4.2(d).

Extension of Results to Spatial Model. Since the key step in the planar case is to find the intersection region inside the two Poincaré cuts, a key difficulty is to determine how to extend this technique to the spatial case. Take as an example the construction of a transit orbit with the itinerary $(X;M,I)$ that goes from the exterior realm to the interior realm of the Jupiter-moon system. Recall that in the spatial case, the unstable manifold tube of the NHIM around L_2 which separates the transit and non-

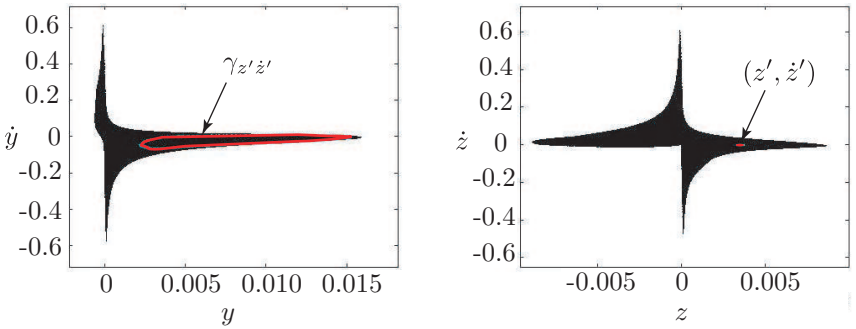


FIGURE 9.4.3. Shown in black are the $y\dot{y}$ (left) and $z\dot{z}$ (right) projections of the 3-dimensional object $\mathcal{C}_1^{+u2}$, the intersection of $W_+^u(\mathcal{M}_h^2)$ with the Poincaré section $x = 1 - \mu$. The set of points in the $y\dot{y}$ projection which approximate a curve, $\gamma_{z'\dot{z}'}$, all have $(z, \dot{z})$ values within the small box shown in the $z\dot{z}$ projection (which appears as a thin strip), centered on $(z', \dot{z}')$. This example is computed in the Jupiter-Europa system for $C = 3.0028$.

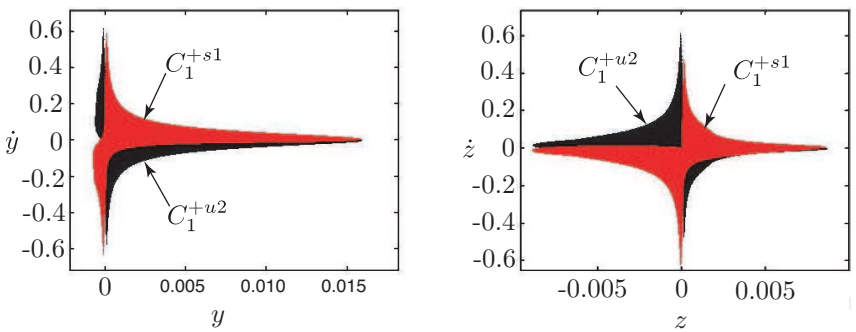


FIGURE 9.4.4. The $y\dot{y}$ (left) and $z\dot{z}$ (right) projections of the 3-dimensional objects $\mathcal{C}_1^{+u2}$ and $\mathcal{C}_1^{+s1}$. This example is computed in the Jupiter-Europa system for $C = 3.0028$.

transit orbits is topologically $S^3 \times \mathbb{R}$. For a transversal cut at $x = 1 - \mu$ (a hyperplane through the moon), the Poincaré cut is a topological 3-sphere S^3 (in $\mathbb{R}^4$). It is not obvious how to find the intersection region inside these two Poincaré cuts (S^3) since both its projections on the $(y, \dot{y})$-plane and the $(z, \dot{z})$-plane are (2-dimensional) disks D^2.

However, in constructing an orbit which transitions from the outside to the inside of a moon's orbit, suppose that we might also want it to have other characteristics above and beyond this gross behavior. We may want to have an orbit which has a particular z-amplitude when it is near the moon. If we set $z = c, \dot{z} = 0$ where c is the desired z-amplitude, the problem of finding the intersection region inside two Poincaré cuts suddenly becomes tractable. Now, the projection of the Poincaré cut of the above unstable

manifold tube on the $(y, \dot{y})$-plane will be a closed curve and any point inside this curve is a $(X; M)$ orbit which has transited from the exterior realm to the moon realm passing through the L_2 equilibrium region. See Figure 9.4.3.

Similarly, we can apply the same techniques to the Poincaré cut of the stable manifold tube to the NHIM around L_1 and find all (M, I) orbits inside a closed curve in the $(y, \dot{y})$-plane. Hence, by using z and $\dot{z}$ as the additional parameters, we can apply the similar techniques that we have developed for the planar case in constructing spatial trajectories with desired itineraries. See Figures 9.4.4 and 9.4.5. What follows is a more detailed description.

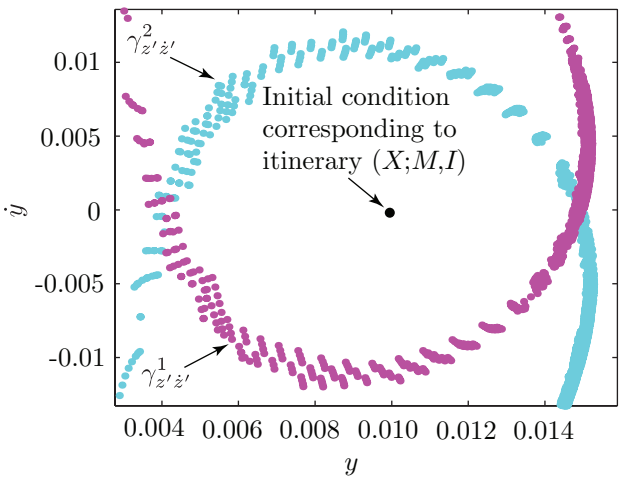


FIGURE 9.4.5. On the $(y, \dot{y})$-plane are shown the points that approximate $\gamma^2_{z'\dot{z}'}$ and $\gamma^1_{z'\dot{z}'}$, the boundaries of $\text{int}(\gamma^2_{z'\dot{z}'})$ and $\text{int}(\gamma^1_{z'\dot{z}'})$, respectively, where $(z', \dot{z}') = (0.0035, 0)$. Note the lemon shaped region of intersection, $\text{int}(\gamma^1_{z'\dot{z}'}) \cap \text{int}(\gamma^2_{z'\dot{z}'})$, in which all orbits have the itinerary $(X; M, I)$. The appearance is similar to Figure 9.4.2(b). The point shown within $\text{int}(\gamma^1_{z'\dot{z}'}) \cap \text{int}(\gamma^2_{z'\dot{z}'})$ is the initial condition for the orbit shown in Figure 9.4.6.

Finding the Poincaré Cuts. We begin with the 15th order normal form expansion near L_1 and L_2. See the appendix of this paper and other references (Gómez, Jorba, Masdemont, and Simó [2001a]; Gómez and Masdemont [2000]; Jorba and Masdemont [1999]). The behavior of orbits in the coordinate system of that normal form, $(q_1, p_1, q_2, p_2, q_3, p_3)$, are qualitatively similar to the behavior of orbits in the linear approximation. This makes the procedure for choosing initial conditions in the L_1 and L_2 equilibrium regions rather simple. In particular, based on our knowledge of the structure for the linear system, we can pick initial conditions which produce a close "shadow" of the stable and unstable manifold tubes ($S^3 \times \mathbb{R}$) asso-

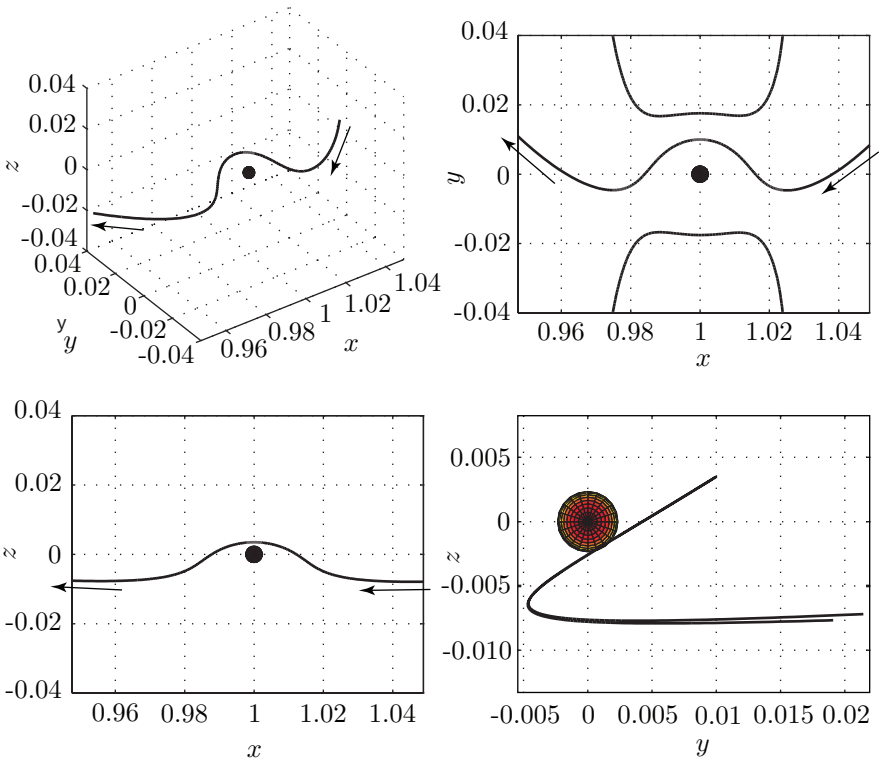


FIGURE 9.4.6. The (X, M, I) transit orbit corresponding to the initial condition in Figure 9.4.5. The orbit is shown in a 3D view and in the three orthographic projections. Europa is shown to scale. The upper right plot includes the $z = 0$ section of the zero velocity surface.

ciated to the normally hyperbolic invariant manifold (NHIM), also called central or neutrally stable manifold, in both the L_1 and L_2 equilibrium regions. As we restrict to an energy surface with energy h, there is only one NHIM per energy surface, denoted $\mathcal{M}_h (\simeq S^3)$.

The initial conditions in $(q_1, p_1, q_2, p_2, q_3, p_3)$ are picked with the qualitative picture of the linear system in mind. The coordinates (q_1, p_1) correspond to the saddle projection, (q_2, p_2) correspond to oscillations within the (x, y) plane, and (q_3, p_3) correspond to oscillations within the z direction. Also note that $q_3 = p_3 = 0$ $(z = \dot{z} = 0)$ corresponds to an invariant manifold of the system, i.e., the planar system is an invariant manifold of the three degree of freedom system.

The initial conditions to approximate the stable and unstable manifolds $(W^s_{\pm}(\mathcal{M}_h), W^u_{\pm}(\mathcal{M}_h))$ are picked via the following procedure. Note that we can be assured that we are obtaining a roughly complete approximation of points along a slice of $W^s_{\pm}(\mathcal{M}_h)$ and $W^u_{\pm}(\mathcal{M}_h)$ since such a slice is compact,

having the structure S^3. Also, we know roughly the picture from the linear case.

1. We fix $q_1 = p_1 = \pm\epsilon$, where ϵ is small. This ensures that almost all of the initial conditions will be for orbits which are transit orbits from one side of the equilibrium region to the other. Specifically $+$ corresponds to right-to-left transit orbits and $-$ corresponds to left-to-right transit orbits. We choose ϵ small so that the initial conditions are near the NHIM $\mathcal{M}_h$ (at $q_1 = p_1 = 0$) and will therefore integrate forward and backward to be near the unstable and stable manifold of $\mathcal{M}_h$, respectively. We choose ϵ to not be too small, or the integrated orbits will take too long to leave the vicinity of $\mathcal{M}_h$.

2. Beginning with $r_v = 0$, and increasing incrementally to some maximum $r_v = r_v^{\max}$, we look for initial conditions with $q_3^2 + p_3^2 = r_v^2$, i.e., along circles in the z oscillation canonical plane. It is reasonable to look along circles centered on the origin $(q_3, p_3) = (0, 0)$ on this canonical plane since the motion is simple harmonic in the linear case and the origin corresponds to an invariant manifold.

3. For each point along the circle, we look for the point on the energy surface in the (q_2, p_2) plane, i.e., the (x, y) oscillation canonical plane. Note, our procedure can tell us if such a point exists and clearly if no point exists, it will not used as an initial condition.

After picking the initial conditions in $(q_1, p_1, q_2, p_2, q_3, p_3)$ coordinates, we transform to the conventional CR3BP coordinates $(x, y, z, \dot{x}, \dot{y}, \dot{z})$ and integrate under the full equations of motion. The integration proceeds until some Poincaré section stopping condition is reached, for example $x = 1-\mu$. We can then use further analysis on the Poincaré section, described below.

Example Itinerary: $(X; M, I)$**.** As an example, suppose we want a transition orbit going from outside to inside the moon's orbit in the Jupiter-moon system. We therefore want right-to-left transit orbits in both the L_1 and L_2 equilibrium regions. Consider the L_2 side. The set of right-to-left transit orbits has the structure $D^4 \times \mathbb{R}$ (where D^4 is a 4-dimensional disk), with boundary $S^3 \times \mathbb{R}$. The boundary is made up of $W_+^s(\mathcal{M}_h^2)$ and $W_+^u(\mathcal{M}_h^2)$, where the $+$ means right-to-left, $\mathcal{M}_h^2$ is the NHIM around L_2 with energy h, and 2 denotes L_2. We pick the initial conditions to approximate $W_+^s(\mathcal{M}_h^2)$ and $W_+^u(\mathcal{M}_h^2)$ as outlined above and then integrate those initial conditions forward in time until they intersect the Poincaré section at $x = 1 - \mu$, a hyperplane passing through the center of the moon.

Since the Hamiltonian energy h (Jacobi constant) is fixed, the set of all values $\mathcal{C} = \{(y, \dot{y}, z, \dot{z})\}$ obtained at the Poincaré section, characterize the branch of the manifold of all libration point orbits around the selected equilibrium point for the particular section. Let us denote the set as $\mathcal{C}_i^{+uj}$, where $+$ denotes the right-to-left branch of the s (stable) or u (unstable)

manifold of the L_j, $j = 1, 2$ libration point orbits at the i-th intersection with $x = 1 - \mu$. We will look at the first intersection, so we have $\mathcal{C}_1^{+u2}$.

The object $\mathcal{C}_1^{+u2}$ is 3-dimensional ($\simeq S^3$) in the 4-dimensional $(y, \dot{y}, z, \dot{z})$ space. For the Jupiter-Europa system, we show $\mathcal{C}_1^{+u2}$ for Jacobi constant $C = 3.0028$ in Figure 9.4.3.

Thus, we suspect that if we pick almost any point $(z', \dot{z}')$ in the $z\dot{z}$ projection, it corresponds to a closed loop $\gamma_{z'\dot{z}'}$ ($\simeq S^1$) in the $y\dot{y}$ projection (see Figure 9.4.3). Any initial condition $(y', \dot{y}', z', \dot{z}')$, where $(y', \dot{y}') \in \gamma_{z'\dot{z}'}$ will be on $W_+^u(\mathcal{M}_h^2)$, and will wind onto a libration point orbit when integrated backward in time. Thus, $\gamma_{z'\dot{z}'}$ defines the boundary of right-to-left transit orbits with $(z, \dot{z}) = (z', \dot{z}')$. If we choose $(y', \dot{y}') \in \text{int}(\gamma_{z'\dot{z}'})$ where $\text{int}(\gamma_{z'\dot{z}'})$ is the region in the $y\dot{y}$ projection enclosed by $\gamma_{z'\dot{z}'}$, then the initial condition $(y', \dot{y}', z', \dot{z}')$ will correspond to a right-to-left transit orbit, which will pass through the L_2 equilibrium region, from the moon realm to outside the moon's orbit, when integrated backward in time.

Similarly, on the L_1 side, we pick the initial conditions to approximate $W_+^s(\mathcal{M}_h^1)$ and $W_+^u(\mathcal{M}_h^1)$ as outlined above and then integrate those initial conditions backward in time until they intersect the Poincaré section at $x = 1 - \mu$, obtaining $\mathcal{C}_1^{+s1}$. We can do a similar construction regarding transit orbits, etc. To distinguish closed loops $\gamma_{z'\dot{z}'}$ from L_1 or L_2, let us call a loop $\gamma_{z'\dot{z}'}^j$ if it is from $L_j, j = 1, 2$.

To find initial conditions for transition orbits which go from outside the moon's orbit to inside the moon's orbit with respect to Jupiter, i.e., orbits which are right-to-left transit orbits in both the L_1 and L_2 equilibrium regions, we need to look at the intersections of the interiors of $\mathcal{C}_1^{+u2}$ and $\mathcal{C}_1^{+s1}$. See Figure 9.4.4.

To find such initial conditions we first look for intersections in the $z\dot{z}$ projection. Consider the projection $\pi_{z\dot{z}} : \mathbb{R}^4 \to \mathbb{R}^2$ given by $(y, \dot{y}, z, \dot{z}) \mapsto (z, \dot{z})$. Consider a point $(y', \dot{y}', z', \dot{z}') \in \pi_{z\dot{z}}(\mathcal{C}_1^{+u2}) \cap \pi_{z\dot{z}}(\mathcal{C}_1^{+s1}) \neq \varnothing$, i.e., a point $(y', \dot{y}', z', \dot{z}')$ where $(z', \dot{z}')$ is in the intersection of the $z\dot{z}$ projections of $\mathcal{C}_1^{+u2}$ and $\mathcal{C}_1^{+s1}$. Transit orbits from outside to inside the moon's orbit are such that $(y', \dot{y}', z', \dot{z}') \in \text{int}(\gamma_{z'\dot{z}'}^1) \cap \text{int}(\gamma_{z'\dot{z}'}^2)$. If $\text{int}(\gamma_{z'\dot{z}'}^1) \cap \text{int}(\gamma_{z'\dot{z}'}^2) = \varnothing$, then no transition exists for that value of $(z', \dot{z}')$. But numerically we find that there are values of $(z', \dot{z}')$ such that $\text{int}(\gamma_{z'\dot{z}'}^1) \cap \text{int}(\gamma_{z'\dot{z}'}^2) \neq \varnothing$. See Figures 9.4.4 and 9.4.5.

In essence we are doing a search for transit orbits by looking at a two parameter set of intersections of the interiors of closed curves, $\gamma_{z\dot{z}}^1$ and $\gamma_{z\dot{z}}^2$ in the $y\dot{y}$ projection, where our two parameters are given by $(z, \dot{z})$. We can reduce this to a *one parameter family of intersections* by restricting to $\dot{z} = 0$. This is a convenient choice since it implies that the orbit is at a critical point (often a maximum or minimum in z when it reaches the surface $x = 1 - \mu$.)

Technically, we are not able to look at curves $\gamma_{z\dot{z}}^j$ belonging to points $(z, \dot{z})$ in the $z\dot{z}$ projection. Since we are approximating the 3-dimensional surface $\mathcal{C}$ by a scattering of points (about a million for the computations

in this paper), we must look not at points $(z, \dot{z})$, but at small boxes $(z \pm \delta z, \dot{z} \pm \delta \dot{z})$ where δz and $\delta \dot{z}$ are small. Since our box in the $z\dot{z}$ projection has a finite size, the points in the $y\dot{y}$ projection corresponding to the points in the box will not all fall on a close curve, but along a slightly broadened curve, a strip, as seen in Figure 9.4.5. For our purposes, we will still refer to the collection of such points as $\gamma^j_{z\dot{z}}$.

9.5 From Ganymede to High Inclination Europa Orbit

Petit Grand Tour. We now apply the techniques we have developed to the construction of a fully three-dimensional Petit Grand Tour of the Jovian moons, extending an earlier planar result (Koon, Lo, Marsden, and Ross [1999, 2002]) We here outline how one systematically constructs a spacecraft tour which begins beyond Ganymede in orbit around Jupiter, makes a close flyby of Ganymede, and finally reaches a high inclination orbit around Europa, consuming less fuel than is possible from standard two-body methods.

Our approach involves the following three key ideas:

1. Treat the Jupiter-Ganymede-Europa-spacecraft four-body problem as two coupled circular restricted 3-body problems. The first 3-body system is the Jupiter-Ganymede-spacecraft system and the second is the Jupiter-Europa-spacecraft system;

2. Use the stable and unstable manifolds of the NHIMs about the Jupiter-Ganymede L_1 and L_2 to find an uncontrolled trajectory from a jovicentric orbit beyond Ganymede to a temporary capture orbit around Ganymede; the spacecraft subsequently leaves Ganymede's vicinity onto a jovicentric orbit interior to Ganymede's orbit;

3. Use the stable manifold of the NHIM around the Jupiter-Europa L_2 to find an uncontrolled trajectory from a jovicentric orbit between Ganymede and Europa to a temporary capture around Europa. Once the spacecraft is temporarily captured around Europa, a propulsion maneuver can be performed when its trajectory is close to Europa (100 km altitude), taking it into a high inclination orbit about the moon. Furthermore, a propulsion maneuver will be needed when transferring from the Jupiter-Ganymede portion of the trajectory to the Jupiter-Europa portion, since the respective transport tubes exist at different energies.

Ganymede to Europa Transfer Mechanism. The construction begins with the patch point, where we connect the Jupiter-Ganymede and Jupiter-Europa portions, and works forward and backward in time toward

each moon's vicinity. The construction is done mainly in the Jupiter-Europa rotating frame using a Poincaré section. After selecting appropriate energies in each 3-body system, respectively, the stable and unstable manifolds of each system's NHIMs are computed. Let ${}^{\text{Gan}}W^u_+(\mathcal{M}^1)$ denote the unstable manifold of Ganymede's L_1 NHIM and ${}^{\text{Eur}}W^s_+(\mathcal{M}^2)$ denote the stable manifold for Europa's L_2 NHIM. We look at the intersection of ${}^{\text{Gan}}W^u_+(\mathcal{M}^1)$ and ${}^{\text{Eur}}W^s_+(\mathcal{M}^2)$ with a common Poincaré section, the surface U_1 in the Jupiter-Europa rotating frame, defined earlier. See Figure 9.5.1.

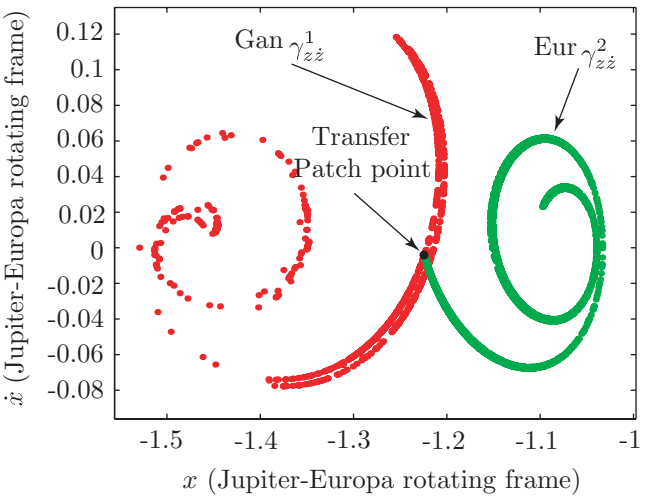


FIGURE 9.5.1. The curves ${}^{\text{Gan}}\gamma^1_{z\dot{z}}$ and ${}^{\text{Eur}}\gamma^2_{z\dot{z}}$ are shown, the intersections of ${}^{\text{Gan}}W^u_+(\mathcal{M}^1)$ and ${}^{\text{Eur}}W^s_+(\mathcal{M}^2)$ with the Poincaré section U_1 in the Jupiter-Europa rotating frame, respectively. Note the small region of intersection, $\text{int}({}^{\text{Gan}}\gamma^1_{z\dot{z}}) \cap \text{int}({}^{\text{Eur}}\gamma^2_{z\dot{z}})$, where the patch point is labeled.

Note that we have the freedom to choose where the Poincaré section is with respect to Ganymede, which determines the relative phases of Europa and Ganymede at the patch point. For simplicity, we select the U_1 surface in the Jupiter-Ganymede rotating frame to coincide with the U_1 surface in the Jupiter-Europa rotating frame at the patch point. Figure 9.5.1 shows the curves ${}^{\text{Gan}}\gamma^1_{z\dot{z}}$ and ${}^{\text{Eur}}\gamma^2_{z\dot{z}}$ on the $(x, \dot{x})$-plane in the Jupiter-Europa rotating frame for all orbits in the Poincaré section with points $(z, \dot{z})$ within $(0.0160 \pm 0.0008, \pm 0.0008)$. The size of this range is about 1000 km in z position and 20 m/s in z velocity.

From Figure 9.5.1, an intersection region on the $x\dot{x}$-projection is seen. We pick a point within this intersection region, but with two differing y velocities; one corresponding to ${}^{\text{Gan}}W^u_+(\mathcal{M}^1)$, the tube of transit orbits coming from Ganymede, and the other corresponding to ${}^{\text{Eur}}W^s_+(\mathcal{M}^2)$, the orbits heading toward Europa. The discrepancy between these two y ve-

locities is the ΔV necessary for a propulsive maneuver to transfer between the two tubes of transit orbits, which exist at different energies.

Four-Body System Approximated by Patched CR3BP. In order to determine the transfer ΔV, we compute the transfer trajectory in the full four-body system, taking into account the gravitational attraction of all three massive bodies on the spacecraft. We use the dynamical channel intersection region in the patched three-body model as an initial guess which we adjust finely to obtain a true four-body bi-circular model trajectory (see Koon, Lo, Marsden, and Ross [2002] for more details).

Figure 1.2.9 is the final end-to-end trajectory. A ΔV of 1214 m/s is required at the location marked. We note that a traditional Hohmann (patched two-body) transfer from Ganymede to Europa requires a ΔV of 2822 m/s. Our value is only 43% of the Hohmann value, which is a substantial savings of on-board fuel. The transfer flight time is about 25 days, well within conceivable mission constraints. This trajectory begins on a jovicentric orbit beyond Ganymede, performs one loop around Ganymede, achieving a close approach of 100 km above the moon's surface. After the transfer between the two moons, a final additional maneuver of 446 m/s is necessary to enter a high inclination (48.6°) circular orbit around Europa at an altitude of 100 km. Thus, the total ΔV for the trajectory is 1660 m/s, still substantially lower than the Hohmann transfer value.

9.6 Normal Form Computations

For the convenience of reader, we have included in the next few sections a brief description of the theoretical basis and the practical steps for developing the software used in the numerical explorations for this chapter.

Linear Behavior and Nonlinear Expansion. In section (9.2), it was shown that the linearized CR3BP Hamiltonian around $L_{1,2}$ has the local behavior of the type saddle×center×center. So, using a real linear and symplectic change of coordinates, it is easy to cast the second order part of the Hamiltonian, H_2 into its real normal form,

$$H_2 = \lambda q_1 p_1 + \frac{\omega_p}{2}(q_2^2 + p_2^2) + \frac{\omega_v}{2}(q_3^2 + p_3^2). \tag{9.6.1}$$

This quadratic normal form has been used in performing qualitative analysis in §9.3 and will be used now as a first step of normal form computations for the full Hamiltonian.

For the following normal form computations it is convenient to "diagonalize" the second-order terms. This is done by introducing the complex

change of coordinates

$$\begin{pmatrix} q_k \\ p_k \end{pmatrix} \to \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} q_k \\ p_k \end{pmatrix}, \qquad k = 2, 3, \tag{9.6.2}$$

and keeping the same variable names for notational convenience. Then, the second-order part of the Hamiltonian becomes

$$H_2 = \lambda q_1 p_1 + i\omega_p q_2 p_2 + i\omega_v q_3 p_3. \tag{9.6.3}$$

In order to have all possible orbits in the center manifold, let us expand the initial Hamiltonian H using the coordinates that give us H_2 as in (9.6.3). Then the expanded Hamiltonian takes the form

$$H(q,p) = H_2(q,p) + \sum_{n\geq 3} H_n(q,p) = H_2(q,p) + \sum_{n\geq 3} h_{ij} q_1^{i_1} p_1^{j_1} q_2^{i_2} p_2^{j_2} q_3^{i_3} p_3^{j_3}, \tag{9.6.4}$$

where H_2 is given in (9.6.3) and H_n denotes an homogeneous polynomial of degree n.

Brief Overview for the Reduction to the Center Manifold. The process of reduction to the center manifold is similar to normal form computations. The objective is to remove some monomials in the expansion of the Hamiltonian, in order to have an invariant manifold tangent to the center directions of H_2. The techniques is through performing canonical changes of variables which can be implemented by means of the Lie series method that is similar to Deprit [1969].

For example, one scheme that has been implemented can eliminate all the monomials with $i_1 \neq j_1$ in (9.6.3), up to an arbitrary order. But in practice, $N = 15$ should be good enough. Although this procedure is not convergent, it produces a very good approximation to the dynamics. Calling $I = q_1 p_1$, and using the inverse of the transformation (9.6.2), the final real Hamiltonian looks like

$$H = H_2(I, q_2, p_2, q_3, p_3) + \sum_{n=3}^{N} H_n(I, q_2, p_2, q_3, p_3) + R_{N+1}(q,p).$$

If we skip the reminder R_{N+1} which is very small near $L_{1,2}$, the truncated Hamiltonian $\bar{H}_N$, which is a polynomial of degree N, has I as a first integral. Setting $I = 0$, we skip the hyperbolic part and reduce the Hamiltonian to the center manifold. This reduction will allow us to compute the center manifold and its stable and unstable manifold in an efficient way.

In the next sections, we will introduce a number of concepts and results in Hamiltonian dynamical systems theory that are indispensable for an understanding of the reduction to the center manifold procedures.

Canonical Transformations. As mentioned earlier, canonical transformations are very useful since they allow one to work on a single function, the Hamiltonian, instead of a system of $2n$ differential equations.

However, to produce a canonical changes of variables is not an easy matter, since it is very difficult to impose the condition that the differential of the change is a symplectic matrix. Fortunately, there exists several techniques to produce such transformations. The one that works for the reduction to the center manifold is based on the following properties of the Hamiltonian flows:

1. Let $\Psi_t(q,p)$ be the time t flow of a Hamiltonian system. Then $(\bar{q},\bar{p}) = \Psi_t(q,p)$ is a canonical transformation.

2. Let $G(q,p)$ be a Hamiltonian system with n degrees of freedom, and let Φ_t be the flow of G. Then the time evolution of a function under the Hamiltonian flow is given by,

$$\frac{d}{dt}(f \circ \Phi_t) = \{f, G\} \circ \Phi_t \tag{9.6.5}$$

for any smooth function f.

The bracket $\{\cdot,\cdot\}$ is called Poisson bracket and is defined for any two smooth functions, $F(q,p)$ and $G(q,p)$, as follow:

$$\{F, G\} = \sum_{i=0}^{3} \left(\frac{\partial F}{\partial q_i}\frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q_i} \right).$$

For notational convenience, $\{F, G\}$ can also be written as $\mathrm{ad}_G(F)$.

Lie Series Method. Now, it is not difficult to prove that to transform a Hamiltonian H by means of the time one flow map of a Hamiltonian G, one can apply the formula

$$\hat{H} \equiv H + \{H, G\} + \frac{1}{2!}\{\{H, G\}, G\} + \frac{1}{3!}\{\{\{H, G\}, G\}, G\} + \cdots, \tag{9.6.6}$$

where $\hat{H}$ denotes the transformed Hamiltonian. This kind of canonical transformations are often referred to as *Lie transform* or Lie series method. The Hamiltonian G is usually called the *generating function* of the transformation (9.6.6). See Giorgilli, Delshams, Fontich, Galgani, and Simó [1989] and references therein for more theoretical details.

Let $F(t) = H \circ \Phi_t$ where Φ_t is the flow of G. Taylor expanding $F(t)$ about $t = 0$ gives

$$F(t) = F(0) + F'(0)t + \frac{1}{2!}F''(0)t^2 + \cdots + \frac{1}{n!}F^{(n)}t^n + \cdots. \tag{9.6.7}$$

Now we want to determine the Taylor coefficients. From the formula (9.6.5), we have,

$$F'(t) = \frac{d}{dt}(H \circ \Phi_t) = \{H, G\} \circ \Phi_t = \mathrm{ad}_G(H) \circ \Phi_t.$$

Differentiate this expression repeatedly, we find

$$\begin{aligned} F''(t) &= \mathrm{ad}_G^2(H) \circ \Phi_t \\ F'''(t) &= \mathrm{ad}_G^3(H) \circ \Phi_t \\ &\vdots \\ F^{(n)}(t) &= \mathrm{ad}_G^n(H) \circ \Phi_t \\ &\vdots \end{aligned}$$

Substitute these expressions with $t = 0$ into (9.6.7), and recall that Φ_0 is an identity map, we obtain the following formula at $t = 1$

$$F(1) = H \circ \Phi_1 = H + \mathrm{ad}_G(H) + \frac{1}{2!}\mathrm{ad}_G^2(H) + \cdots + \frac{1}{n!}\mathrm{ad}_G^n(H) + \cdots .$$

which proves the result (9.6.6) as claimed.

This expression is very suitable for effective computations, since it can be implemented in a computer. Note that all the operations involved are very simple if we are working with some kind of expansions such as power series expansions, Fourier expansions, etc. One can argue that the problem for this kind of transformation from a practical point of view is that it is defined by an infinite series. This is not a difficulty since we usually work with a finite truncation of these series which produces a high order approximation to the desired results that, in many cases, are good enough for practical purposes. On the other hand, it is possible to derive rigorous estimates on the size of the reminder so one can obtain bounds on the error of the results obtained with the truncated series. See Giorgilli, Delshams, Fontich, Galgani, and Simó [1989] and Simó [1989] for more information.

Elimination of Third-Order Terms. Note that if P and Q are two homogeneous polynomials of degree r and s respectively, then $\{P, Q\}$ is an homogeneous polynomial of degree $r + s - 2$. This means that if G_3 is an homogeneous polynomial of degree 3 used as a generating function, then the homogeneous polynomials of degree n, $\hat{H}_n$, such that $\hat{H} = \sum_{n\geq 2} \hat{H}_n$ are given by,

$$\begin{aligned} \hat{H}_2 &= H_2, \\ \hat{H}_3 &= H_3 + \{H_2, G_3\}, \\ \hat{H}_4 &= H_4 + \{H_3, G_3\} + \frac{1}{2!}\{\{H_2, G_3\}, G_3\}, \\ &\cdots\cdots\cdots \end{aligned}$$

If we are interested in removing all the terms of order three in the new Hamiltonian, i.e., to have $\hat{H}_3 = 0$, we must choose G_3 such that it solves the homological equation $\{H_2, G_3\} = -H_3$.

For notation convenience, we will use the following notation below. If $x = (x_1, \ldots, x_m)$ is a vector of complex numbers and $k = (k_1, \ldots, k_m)$ is an integer vector, we denote by x^k the term $x_1^{k_1} \cdots x_n^{k_m}$ (in this context we define 0^0 as 1). Moreover, we define $|k|$ as $\sum_j |k_j|$.

Let us denote

$$H_3(q,p) = \sum_{|k_q|+|k_p|=3} h_{k_q,k_p} q^{k_q} p^{k_p}, \quad G_3(q,p) = \sum_{|k_q|+|k_p|=3} g_{k_q,k_p} q^{k_q} p^{k_p},$$

and $H_2(q,p) = \sum_{j=1}^{3} \omega_j q_j p_j$, where $\omega_1 = \lambda$, $\omega_2 = i\omega_p$ and $\omega_3 = i\omega_v$.

As

$$\{H_2, G_3\} = \sum_{|k_q|+|k_p|=3} \langle k_p - k_q, \omega \rangle g_{k_q,k_p} q^{k_q} p^{k_p}, \quad \omega = (\omega_1, \omega_2, \omega_3),$$

it is immediate to obtain

$$G_3(q,p) = \sum_{|k_q|+|k_p|=3} \frac{-h_{k_q,k_p}}{\langle k_p - k_q, \omega \rangle} q^{k_q} p^{k_p},$$

that is well defined because the condition $|k_q| + |k_p| = 3$ implies that $\langle k_p - k_q, \omega \rangle$ is different from zero. Note that G_3 is so easily obtained because of the "diagonal" form of H_2 given in (9.6.3).

This procedure can be used recurrently trying to find an homogeneous polynomial of degree four, G_4, to remove terms of order four of the new Hamiltonian, $\hat{H}$, and so on. Nevertheless, we must point out that some resonant terms, even of order four, cannot be canceled. Anyway, this process is used to compute what is known as the Birkhoff normal form of the Hamiltonian which has the minimum number of monomials up to some degree.

9.7 Reduction to the Center Manifold

Although the reduction to the center manifold is based on this scheme, we only need to remove the instability associated with hyperbolic character of the Hamiltonian H. Note that the second-order part of the Hamiltonian H_2 yields the linear part of the Hamiltonian equations, and so, the instability is associated with the term $\lambda q_1 p_1$. For this linear approximation of the Hamiltonian equations, the center part can be obtained by setting $q_1 = p_1 = 0$. If we want the trajectory remains tangent to this space (i.e., having $q_1(t) = p_1(t) = 0$ for all $t > 0$) when adding the nonlinear terms, we need

to have $\dot{q}_1(0) = \dot{p}_1(0) = 0$ once we have set $q_1(0) = p_1(0) = 0$. Then, because of the autonomous character of the Hamiltonian system, we will obtain $q_1(t) = p_1(t) = 0$ for all $t \geq 0$.

Recalling that the Hamiltonian equations associated with a Hamiltonian $H(q, p)$ are given by,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

In particular,

$$\dot{q}_1 = \frac{\partial H}{\partial p_1} = \lambda q_1 + \sum_{n\geq 3} h_{ij} q_1^{i_1} p_1^{j_1 - 1} q_2^{i_2} p_2^{j_2} q_3^{i_3} p_3^{j_3}$$

$$\dot{p}_1 = -\frac{\partial H}{\partial q_1} = -\lambda p_1 - \sum_{n\geq 3} h_{ij} q_1^{i_1 - 1} p_1^{j_1} q_2^{i_2} p_2^{j_2} q_3^{i_3} p_3^{j_3}.$$

One can get the required condition, $\dot{q}_1(0) = \dot{p}_1(0) = 0$ when $q_1(0) = p_1(0) = 0$, if in the series expansion of the Hamiltonian H, all the monomials, $h_{ij} q^i p^j$, with $i_1 + j_1 = 1$, have $h_{ij} = 0$. This happens if there are no monomials with $i_1 + j_1 = 1$. Since this *minimalist* expansion needs to cancel least monomials in (9.6.4), in principle, it may be better behaved both in terms of convergence and from a numerical point of view. Of course other expansions could give us the same required tangency, such as the one which eliminates all the monomials with $i_1 \neq j_1$.

Both expansions have been implemented by symbolic manipulators that can carry out the procedure up to an arbitrary order. For practical purposes, and in order to have an acceptable balance precision and time computing requirements, the reduction to the center manifold scheme has been carried out up to order $N = 15$.

Approximate First Integral. In order to produce an approximate first integral having the center manifold as a level surface, it is enough to eliminate the monomials where $i_1 \neq j_1$.

This implies that the generating function G_3 is

$$G_3(q, p) = \sum_{n=3} \frac{-h_{ij}}{(j_1 - i_1)\lambda + (j_2 - i_2) i\omega_p + (j_3 - i_3) i\omega_v} q_1^{i_1} p_1^{j_1} q_2^{i_2} p_2^{j_2} q_3^{i_3} p_3^{j_3}, \tag{9.7.1}$$

where $i_1 \neq j_1$. This follows immediately from

$$\begin{aligned} \langle k_p - k_q, \omega \rangle &= (j_1 - i_1)\omega_1 + (j_2 - i_1)\omega_2 + (j_3 - i_3)\omega_3 \\ &= (j_1 - i_1)\lambda + (j_2 - i_1) i\omega_p + (j_3 - i_3) i\omega_v, \end{aligned}$$

Then, the transformed Hamiltonian $\hat{H}$ takes the form

$$\hat{H}(q, p) = H_2(q, p) + \hat{H}_3(q, p) + \hat{H}_4(q, p) + \cdots, \tag{9.7.2}$$

where $\hat{H}_3(q,p) \equiv \hat{H}_3(q_1p_1, q_2, p_2, q_3, p_3)$, Note that $\hat{H}_3$ depends on the product q_1p_1, not on each variable separately.

This process can be carried out up to any finite order n and the generating function used to remove monomials of degree n will be of the form

$$G_n(q,p) = \sum_{|i|+|j|=n} \frac{-h_{ij}}{(j_1-i_1)\lambda + (j_2-i_2)i\omega_p + (j_3-i_3)i\omega_v} q_1^{i_1} p_1^{j_1} q_2^{i_2} p_2^{j_2} q_3^{i_3} p_3^{j_3}, \tag{9.7.3}$$

where $i_1 \neq j_1$.

It is interesting to note the absence of small divisors during the entire process. Since we only want to eliminate all monomials with $i_1 \neq j_1$, the real part of the the denominators are non-zero. So, the denominators that appear in the generating functions (9.7.3) are never zero and can be bounded from below

$$|(j_1-i_1)\lambda + (j_2-i_2)i\omega_p + (j_3-i_3)i\omega_v| \geq |\lambda|.$$

For this reason, the divergence of this process is very mild and has been observed when this process is stopped at some degree N. Hence, the remainder is very small in quite a big neighborhood of the equilibrium point.

After this process has been carried out, up to a finite order N, we obtain a Hamiltonian of the form

$$\bar{H}(q,p) = \bar{H}_N(q,p) + R_N(q,p),$$

where $H_N(q,p) \equiv H_N(q_1p_1, q_2, p_2, q_3, p_3)$ is a polynomial of degree N and R_N is a remainder of order greater then N. Note that H_N depends on the product q_1p_1 while the remainder depends on the two variables q_1 and p_1 separately.

Now, neglect the remainder and apply the canonical change given by $I = q_1p_1$, we obtain the Hamiltonian $\bar{H}_N(I, q_2, p_2, q_3, p_3)$ that has I_1 as a first integral. Finally, use the inverse change of variables of (9.6.2), the truncated Hamiltonian $\bar{H}_N$ can be expanded in real form

$$\bar{H}_N(q,p) = H_2(q,p) + \sum_{n=3}^{N} H_n(q,p),$$

where explicitly,

$$H_2(q,p) = H_2(q_1,p_1,q_2,p_2,q_3,p_3) = \lambda q_1p_2 + \frac{\omega_p}{2}\left(q_2^2+p_2^2\right) + \frac{\omega_v}{2}\left(q_3^2+p_3^2\right).$$

For convenience, the variables are called again q, p.

Explicit Formula for Change of Variables. An explicit expression for the change of variables that goes from the coordinates of the reduction to the center manifold to the coordinates corresponding to Hamiltonian

(9.6.4) can be obtained in the following way: once the generating function G_3 has been obtained, we can compute

$$\tilde{q}_j = q_j + \{q_j, G_3\} + \frac{1}{2!}\{\{q_j, G_3\}, G_3\} + \frac{1}{3!}\{\{\{q_j, G_3\}, G_3\}, G_3\} + \cdots, \tag{9.7.4}$$

$$\tilde{p}_j = p_j + \{p_j, G_3\} + \frac{1}{2!}\{\{p_j, G_3\}, G_3\} + \frac{1}{3!}\{\{\{p_j, G_3\}, G_3\}, G_3\} + \cdots, \tag{9.7.5}$$

that produces the transformation that sends the coordinates of (9.6.4), given by the variables $(\tilde{q}, \tilde{p})$ into the coordinates of (9.7.2), represented by the variables (q, p). In the next step, the generating function G_4 is applied to the right-hand side of equations (9.7.4) and (9.7.5), to obtain the change corresponding to fourth order, and so on. Finally, these expansions are put into real form in the same way as the Hamiltonian.

From now on, and to simplify notation, the final variables will still be called q and p.

After all these changes of variables, the initial complexified Hamiltonian around the collinear libration points has been expanded in the following form

$$H(q, p) = \bar{H}_N(q, p) + R_N(q, p),$$

where $\bar{H}_N(q, p)$ is a polynomial of degree N without terms of $i_1 + j_1 = 1$ in the minimalist case, or without terms of $i_1 \neq j_1$ in the first integral case. $R_N(q, p)$ is a remainder of order $N + 1$ which is very small near $L_{1,2}$ and will be skipped in further computations.

9.8 NHIM and its Stable and Unstable Manifolds

Then, setting $I_1 = 0$ (this is to skip the hyperbolic behavior) we obtain a two degrees of freedom Hamiltonian, $\bar{H}_N(0, \bar{q}, \bar{p})$, $\bar{q} = (q_2, q_3)$, $\bar{p} = (p_2, p_3)$, that represents (up to some finite order N) the dynamics inside the center manifold. Finally, the Hamiltonian is put into its real form by using the inverse of (9.6.2).

As discussed previously, in the case that all the monomial in H_n with $i_1 \neq j_1$ has been eliminated, the truncated Hamiltonian $\bar{H}_N$ has a first integral, $I = q_1 p_1$. This is because $\bar{H}_N$ is given by

$$\bar{H}_N = H_2(I, q_2, p_2, q_3, p_3) + \sum_{n=3}^{N} H_n(I, q_2, p_2, q_3, p_3).$$

Let f be a function of the center manifold variables (q_2, p_2, q_3, p_3) defined as follow

$$f(q_2, p_2, q_3, p_3) = \sum_{n=3}^{N} H_n(0, q_2, p_2, q_3, p_3).$$

Then, f is at least of third order. Notice that the invariant manifold $\mathcal{M}_h^3$ defined by

$$\mathcal{M}_h^3 = \left\{(q,p) \mid \frac{\nu}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega}{2}\left(q_3^2 + p_3^2\right) + f(q_2, p_2, q_3, p_3) = h, \;\; q_1 = p_1 = 0\right\}$$

is the normally hyperbolic invariant manifold (NHIM) for the nonlinear system which corresponds to the 3-sphere (9.3.1) for the linearized system. In a small neigborhood of the equilibrium point, since the nonlinear terms are much smaller than the linear terms, the 3-sphere for the linear problem becomes a deformed sphere for the nonlinear problem. Moreover, since NHIMs persist under perturbation, this deformed sphere $\mathcal{M}_h^3$ still has stable and unstable manifolds which are given by

$$W_{\pm}^s(\mathcal{M}_h^3) = \left\{(q,p) \mid \frac{\nu}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega}{2}\left(q_3^2 + p_3^2\right) + f(q_2, p_2, q_3, p_3) = h, q_1 = 0\right\}$$

$$W_{\pm}^u(\mathcal{M}_h^3) = \left\{(q,p) \mid \frac{\nu}{2}\left(q_2^2 + p_2^2\right) + \frac{\omega}{2}\left(q_3^2 + p_3^2\right) + f(q_2, p_2, q_3, p_3) = h, p_1 = 0\right\}.$$

Notice the similarity between the formulas above and those for the linearized problem (9.3.2 and 9.3.3), especially given the fact that these two coordinate systems are linked by a near-identity transformation.

10

Multi-Moon Orbiters and Low Thrust Trajectories

In this chapter we discuss some of the latest research topics related to the subtleties of the 3- and N-body problems as related to mission trajectory design. Some of the ideas are only outlined in general terms, and details are left in the references.

In the first part of the chapter (§10.1-10.5) we discuss *multi-moon orbiters*, an extension of the work in Chapters 5 and 9 to a class of low energy missions in which a single spacecraft orbits multiple moons of a planet. This concept achieves fuel savings by exploiting not only tube dynamics, but also *lobe dynamics* to increase fuel savings tremendously. Far from being simply a curiousity, the multi-moon orbiter (MMO) is one of the possible orbit structures for missions to explore the moon systems of the outer solar system, such as a tour of Jupiter's icy moons.

In the second part of the chapter (§10.6) we discuss *low thrust trajectories*, another area of emerging interest. For low thrust trajectories the propulsion delivers a low level of continuous thrust, acting effectively as a small perturbation on the natural dynamics. Using this small continuous control efficiently will have important implications for multi-moon orbiter and other missions in the future which intend to use low thrust technology. Low thrust trajectories, unlike the high thrust trajectories we have discussed in previous chapters, are trajectories which, while the thrust is on, continuously change the spacecraft's three-body energy, eq. (2.3.13). One thus can imagine a picture where the Hill's region changes in time, moving between the cases depicted in Figure 2.4.2.

Open problems regarding both MMOs and low thrust trajectory design are mentioned.

10.1 Introduction to Multi-Moon Orbiter Design

Mission to Icy Moons of Jupiter is Likely. There has been much recent interest in sending a spacecraft to perform extended observations of several of Jupiter's moons. Europa is thought to be a place hospitable to life because of the vast, liquid oceans that may exist under its icy crust. Two other Jupiter moons, Ganymede and Callisto, are now also thought to have liquid water beneath their surfaces. A proposed mission to Europa, and perhaps also Ganymede and Callisto, would attempt to map these regions of liquid water for follow-on missions. The recent discovery of life in the ice of Lake Vostok, a lake deep beneath the Antarctic ice cap, lends impetus to an icy moon mission with the suggestion that life may be possible on such worlds (Karl, Bird, Bjorkman, Houlihan, Shackelford, and Tupas [1999]). A mission to tour the moons of Jupiter may likely be undertaken in the near future.

Multi-Moon Orbiter: Orbiting Several Moons with One Spacecraft. In response to the scientific interest in Jupiter's moons, the MMO tour concept was introduced, wherein a single sophisticated spacecraft is capable of jumping from an orbit around one jovian moon to an orbit around another (Ross, Koon, Lo, and Marsden [2003]). This would allow long duration observations of each moon, compared to brief flybys. The MMO embodies a radical departure from the past four decades of planetary exploration. Such a capability allows close, detailed, and long-lerm studies to be made of many of the members of Jupiter's retinue of 40 (or more) moons. Furthermore, the ΔV requirements for such a mission can be very low if the technique of *low energy inter-moon transfer via resonant gravity assists* is used. In the phase space, the resonant gravity assists (GAs) can be understood in the framework of lobe dynamics, described in the following sections.

As an example, by using small impulsive thrusts totaling only 22 m/s, a spacecraft initially injected into a jovian orbit can be directed into an inclined, elliptical capture orbit around Europa. Enroute, the spacecraft orbits both Callisto and Ganymede for long duration using the ballistic capture and escape methodology described in previous chapters. This example tour is shown in Figure 1.2.10 in Chapter 1.

The MMO, constructed using the patched three-body approach (P3BA) discussed in Chapters 5 and 9 augmented by resonant gravity assists, should enhance existing trajectory design capabilities. The techniques used to design a MMO are very general. The same techniques may be applied to a

broad range of multi-body missions from the Earth's Neighborhood to other regions of the solar system, e.g., a tour of the moons of Uranus (Topputo, Massimiliano, and Finzi [2004]).

An Extension of the Petit Grand Tour. The MMO tour grew out of the method introduced in in §9.5 for the Petit Grand Tour. In the first part of this chapter, the previous method is extended significantly, introducing the use of multiple resonant GAs. This method was inspired by the work of Sweetser and others at JPL who designed the first nominal trajectory for a Europa orbiter mission (Sweetser, Maddock, Johannesen, Bell, Penzo, Wolf, Williams, Matousek, and Weinstein [1997]). But the current method is an improvement upon previous trajectories in that the ΔV requirements are greatly reduced by including special sequences of GAs with the jovian moons. One can find tours with a deterministic ΔV as low as $\sim$20 m/s as compared with $\sim$1500 m/s using previous methods (Ludwinski, Guman, Johannesen, Mitchell, and Staehle [1998]). In fact, a ΔV of $\sim$20 m/s is on the order of statistical navigation errors, and is therefore exceptionally low.

The resonant GAs with the moons can be viewed as follows. By performing small maneuvers to achieve a particular moon/spacecraft geometry at close approach, the spacecraft can jump between mean motion resonances with the moon. The "jumping" provides an effective "ΔV kick", and thereby the deterministic propulsive ΔV required for the MMO trajectory is very low. The problem of finding the "jumping" trajectories can be reduced to a problem of finding points on a Poincaré section which move between "resonance regions."

Generating a Family of Trajectories. Our basic goal in the first part of the chapter is not to construct flight-ready end-to-end MMO trajectories, but rather to introduce the building blocks of the method. Our motivating application will be the construction of a family of inter-moon transfers between moons of Jupiter. As each transfer requires a particular ΔV and time of flight (TOF), the family can be represented as a set of points in the (TOF, ΔV) plane. Therefore, the construction of such a family illustrates the trade-off between fuel consumption and TOF, which has been observed for low energy MMO trajectories (Ross, Koon, Lo, and Marsden [2004]). The dramatically low ΔV needed for the tour shown in Figure 1.2.10 is achieved at the expense of time—it has a TOF of about four years, mostly spent in the inter-moon transfer phase. Although this is likely too long to be acceptable for an actual mission, an exploration of the ΔV vs. TOF trade-off for inter-moon transfers, as done in §10.4, reveals how time savings can be achieved for low cost in additional fuel fuel. In fact, we find a roughly linear relationship between ΔV vs. TOF, and conclude that a reasonable TOF for a MMO can be achieved using a feasible ΔV.

The Building Blocks for MMO Trajectory Generation. Following the point of view taken in this book, we use simple models which provide

dynamical insight for MMO trajectory generation and are computationally tractable. The forward-backward method in the restricted three-body problem phase space is used (Schroer and Ott [1997]; Koon, Lo, Marsden, and Ross [1999]). The influence of only one moon at a time is considered. Criteria are established for determining when the switch from one moon's influence to another occurs.

The P3BA discussed in Chapter 9 considers the motion of a particle (or spacecraft, if controls are permitted) in the field of n bodies, considered two at a time, e.g., Jupiter and its ith moon, M_i. When the trajectory of a spacecraft comes close to the orbit of M_i, the perturbation of the spacecraft's motion away from purely Keplerian motion about Jupiter is dominated by M_i. In this situation, we say that the spacecraft's motion is well modeled by the Jupiter-M_i-spacecraft restricted three-body problem. Within the three-body problem, we can take advantage of phase space structures such as tubes of capture and escape, as well as lobes associated with movement between orbital resonances. Both tubes and lobes, specifically the dynamics associated with them, are important for the design of a MMO trajectory.

The design of a MMO of the jovian system is guided by four main ideas:

1. The multi-moon transfer problem can be broken down into sequential transfers between two moons only. For sake of discussion let us consider the moons Ganymede and Europa.

2. The motion of the spacecraft in the gravitational field of the three bodies Jupiter, Ganymede, and Europa is approximated by two segments of purely three body motion in the circular, restricted three-body model. The trajectory segment in the first three body system, Jupiter-Ganymede-spacecraft, is appropriately patched to the segment in the Jupiter-Europa-spacecraft three-body system.

3. For each segment of purely three body motion, the invariant manifolds tubes of L_1 and L_2 bound orbits (including periodic orbits) leading toward or away from temporary capture around a moon are used to construct an orbit with the desired behaviors. Portions of these tubes are "carried" by the lobes mediating movement between orbital resonances. Directed movement between orbital resonances is what allows a spacecraft to achieve large changes in its orbit. When the spacecraft's motion, as modeled in one three-body system, reaches an orbit whereby it can switch to another three-body system, we switch or "patch" the three-body model to the new system.

4. This initial guess solution is then refined to obtain a trajectory in a more accurate four-body model. Evidence suggests that these initial guesses are very good, even in the full N-body model and considering the orbital eccentricity of the moons (Yamato and Spencer [2003])

Inter-Moon Transfer and the Switching Orbit. During the inter-moon transfer—where one wants to leave a moon and transfer to another moon, closer in to Jupiter—we consider the transfer in two portions, shown schematically in Figure 10.1.1, with M_1 as the outer moon and M_2 as

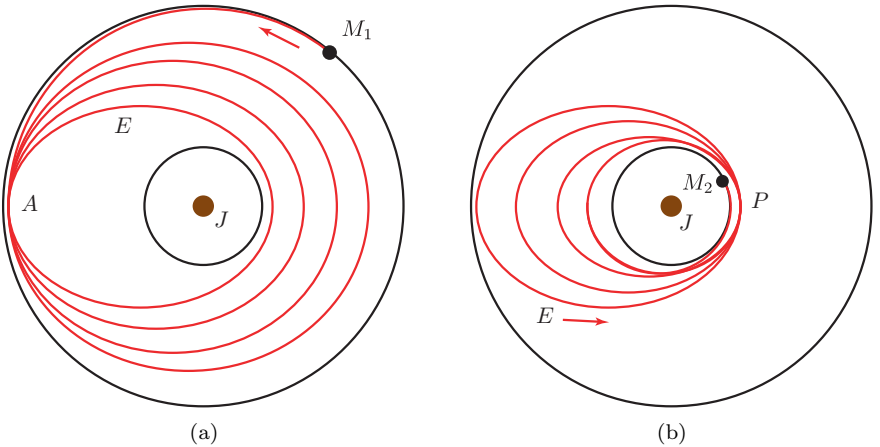


FIGURE 10.1.1. **Inter-moon transfer via resonant gravity assists.** (a) The orbits of two Jovian moons are shown as circles. Upon exiting the outer moon's (M_1's) sphere-of-influence, the spacecraft proceeds under third body effects onto an elliptical orbit about Jupiter. The spacecraft gets a gravity assist from the outer moon when it passes through apojove (denoted A). The several close encounters with M_1 occur at roughly the same Jupiter-moon-spacecraft angle. Once the spacecraft orbit comes close to grazing the orbit of the inner moon, M_2 (in fact, grazing the orbit of M_2's L_2 point), the inner moon becomes the dominant perturber. The spacecraft orbit where this occurs is denoted E. (b) The spacecraft now receives gravity assists from M_2 at perijove (P), where the near-resonance condition also applies. The spacecraft is then ballistically captured by M_2.

the inner moon. In the first portion, the transfer determination problem becomes one of finding an appropriate solution of the Jupiter-M_1-spacecraft problem which jumps between orbital resonances with M_1, i.e., performing resonant GAs to decrease the perijove. M_1's perturbation is only significant over a small portion of the spacecraft trajectory near apojove (A in Figure 10.1.1(a)). The effect of M_1 is to impart an impulse to the spacecraft, effectively imparting a ΔV.

The perijove is decreased by repeated GAs by M_1 (of the appropriate geometry) until it has a value close to M_2's orbit, or more correctly, close to the orbit of M_2's L_2. We can then assume that a GA can be achieved with M_2 with an appropriate geometry such that M_2 becomes the dominant perturber and all subsequent GAs will be with M_2 only. The problem of finding GAs of the "appropriate geometry" can be solved with the help of lobe dynamics, discussed in §10.3.

The arc of the spacecraft's trajectory at which the spacecraft's perturbation switches from being dominated by moon M_1 to being dominated by M_2 is called the *switching orbit*. A rocket burn maneuver need not be necessary to effect this switch. The set of possible switching orbits is the *switching region* of the P3BA. It is the analogue of the "sphere of influence" concept used in the patched conic approximation, which guides a mission designer regarding when to switch the central body for the model of the spacecraft's Keplerian motion.

The spacecraft orbit where M_2 takes over as the perturbing moon is denoted E in Figures 10.1.1(a) and 10.1.1(b). The spacecraft now gets GA's from the inner moon at perijove (P). One can then search for solutions of the Jupiter-M_2-spacecraft problem which cause the apojove to decrease at every close encounter with M_2, causing the spacecraft's orbit to get more and more circular, as in Figure 10.1.1(b). When a particular resonance is reached (see, e.g., Figure 2.3.3), the spacecraft can then be ballistically captured by the inner moon at M_2. We note that a similar phenomenon has been observed in studies of Earth to lunar transfer trajectories (Schroer and Ott [1997]; Schoenmaekers, Horas, and Pulido [2001]), and in studies of natural minor body motion in the solar system (Ross [2003, 2004]; Ćuk and Burns [2004]).

10.2 Inter-Moon Transfer Via the Resonant Structure of Phase Space

Solutions to the four-body problem which lead to the behavior shown schematically in Figure 10.1.1 have been found numerically and the phenomena can be partially explained in terms of the P3BA (Ross, Koon, Lo, and Marsden [2003, 2004]). To switch between neighboring pairs of moons, a spacecraft must traverse several subregions of the three-body problem phase space. Looking at the four-body problem as two three-body problems, the subregions are understood as *resonance regions*, where the resonance is between the spacecraft orbital period and the dominant moon's orbit period around Jupiter, respectively. Such a resonance is referred to as a *mean motion resonance*.

We can visualize the resonances on a Poincaré section. For example, in the m_1-m_2-spacecraft system, we might choose a section in the exterior realm as shown in Figure 10.2.1(a). Such a procedure reveals a phase space, as in Figure 10.2.1(b), consisting of overlapping resonance regions (Malhotra [1996]; Schroer and Ott [1997]). This means that movement amongst resonances is possible.

The points in Figure 10.2.1(b) were computed on a Poincaré section which is area preserving. Following the convention of Malhotra [1996] and

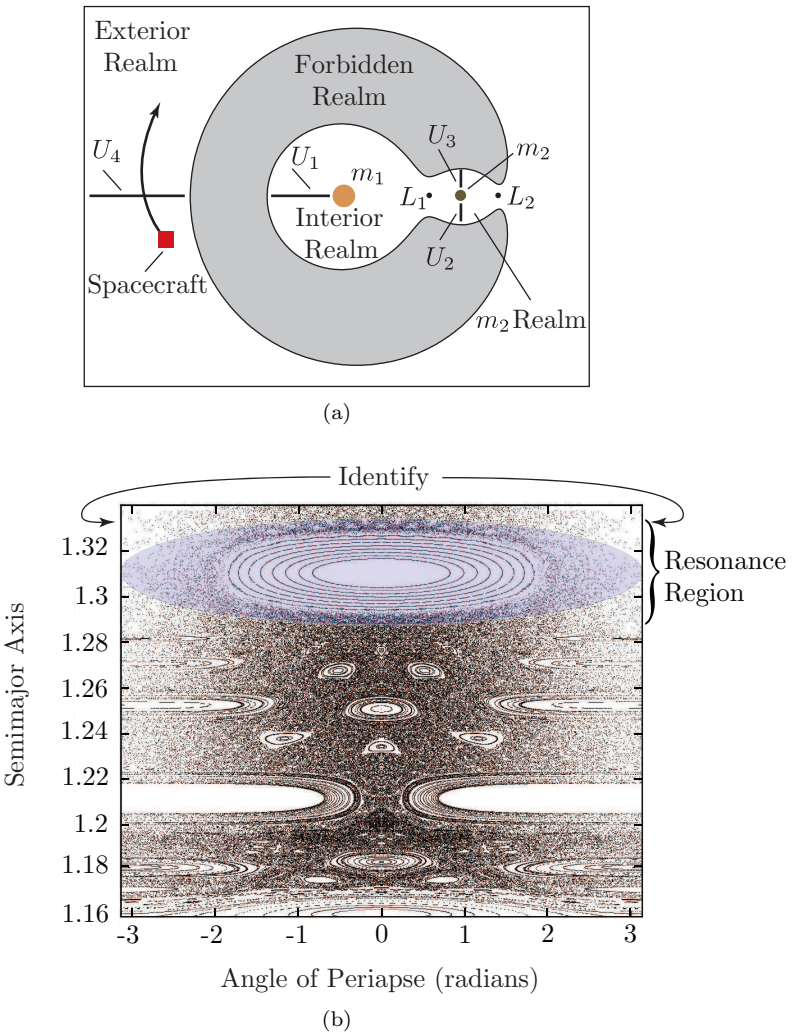


FIGURE 10.2.1. **A Poincaré section showing the resonances.** (a) The location of the four Poincaré sections introduced in Chapter 3 are shown—U_1, U_2, U_3, and U_4. (b) The resonant structure on the U_4 section is shown. The resonance regions are the roughly horizontal bands, one of which is highlighted. Where the resonance region pinches to a point there is a saddle point associated with the resonance, corresponding to an unstable resonant orbit in the full phase space. Stable mean motion resonances inside the resonance regions appear as target patterns.

Koon, Lo, Marsden, and Ross [2000], the points were transformed into quasi-action angle coordinates which give physical insight and allow ease of visualization, but are not area preserving. The x-axis is an angular coordinate and thus the right and left sides are to be identified. The y-axis

is the semimajor axis. The coordinate system is related to the *Delaunay* variables described in the Supplement below.

From Figure 10.2.1(b), a wealth of information can be gained. The phase space is 'unpeeled' to reveal the resonance structure which governs transport in the semimajor axis direction, i.e., from one Keplerian orbit to another. To go from, say, a large to a small semimajor axis (i.e., the top to the bottom of Figure 10.2.1(b)), several resonance regions must be traversed. We can see the locations and widths of the stable resonance libration zones (the target patterns) which are bounded by chaotic layers. Furthermore, the transport around the stable resonance zones is determined by additional structure within the chaotic layers, not seen here. The additional structure is related to saddle points in the chaotic layers and the transport is explained using lobe dynamics below.

Lobe Dynamics Reveals Fastest Resonance Traversing Trajectories. Lobe dynamics (Vetells us the most important spacecraft trajectories, i.e., the uncontrolled trajectories which traverse the resonance regions in the shortest time. Consequently, it proves useful for designing low energy spacecraft trajectories, such as shown in Figure 10.2.2(a). We show an initial condition in the upper right hand side of this Poincaré section which emerges from a tube. This is a trajectory which may have come from the realm around a moon like Ganymede. The spacecraft moves through the phase space as revealed in the Poincaré section, jumping between resonance regions under the natural dynamics of the three-body problem, i.e., at zero fuel cost. The spacecraft arrives at the tube for a destination moon like Europa in the lower left hand side. Figure 10.2.2(b) shows a schematic of the corresponding trajectory in inertial space. This provides us with the pieces for switching from one moon to another as shown in Figure 10.1.1.

Supplement: Delaunay Variables

Delaunay Variables: Inertial Frame. The PCR3BP is a perturbation of the two-body Keplerian problem. We can re-write the Hamiltonian (2.3.11) in the inertial frame as

$$H = H_0 + \mu H_1, \tag{10.2.1}$$

where H_0 is the Keplerian part and μH_1 the perturbing part. $H_0 = -\frac{1}{2a}$ with a the instantaneous semimajor axis of the particle's orbit around the m_1-m_2 center of mass (Wisdom [1980]). In terms of the *Delaunay variables* or *Delaunay canonical elements* (see, e.g., Delaunay [1860]; Szebehely [1967]; Chang and Marsden [2003]), $H_0 = -\frac{1}{2L^2}$ and $H_1 = -R$, where

$$R = \sum_{i=0}^{\infty} \sum_{j=-\infty}^{\infty} K^{(i,j)} \cos N^{(i,j)}, \tag{10.2.2}$$

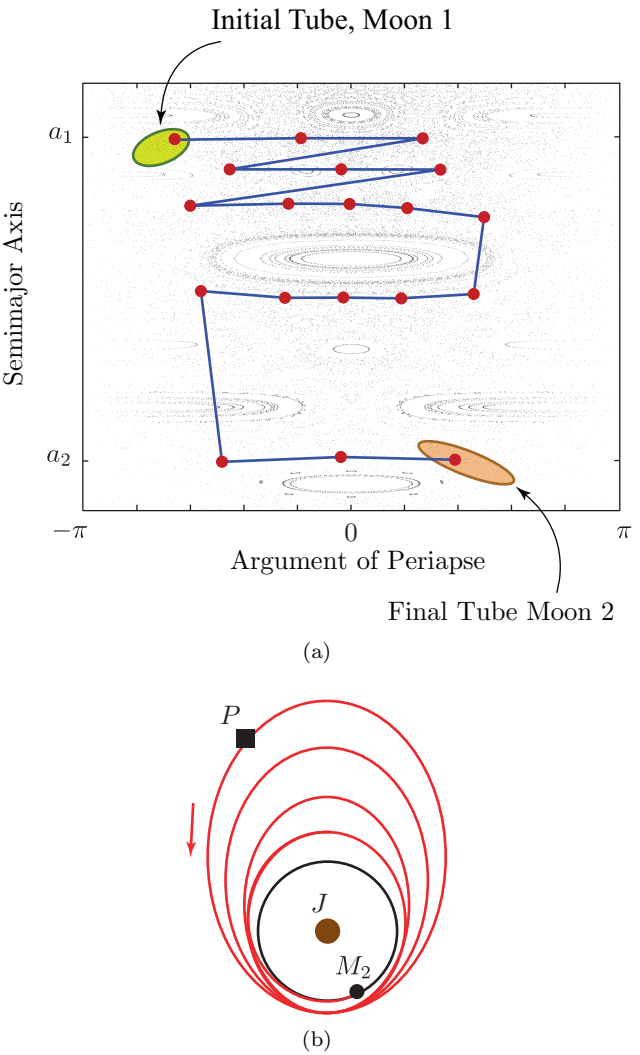


FIGURE 10.2.2. **Jumping between resonance regions leads to large orbit changes at zero cost.** (a) As shown in this Poincaré section, a spacecraft emerges from a tube coming from moon 1 in an orbit with semimajor axis of a_1 around Jupiter. The spacecraft moves through the phase space as shown, jumping between resonance regions until it reaches a smaller semimajor axis, a_2, where it enters the tube going to another moon. In (b), a schematic of the corresponding trajectory of the spacecraft P in inertial space is shown along with Jupiter (J) and the destination moon (M_2).

and

$$N^{(i,j)} = il + j(t - g). \tag{10.2.3}$$

Notice that this Hamiltonian is time dependent.

In terms of the traditional osculating elliptic elements, the canonical momenta are $L = \sqrt{a}$ and $G = L\sqrt{1-e^2}$, where a is the semimajor axis and e is the eccentricity. $K^{(i,j)}$ in the above expression is a function of L and G (see Wisdom [1980]). We note that G is the angular momentum of the particle's orbit around the m_1-m_2 center of mass. The coordinates conjugate to L and G are the mean anomaly l and the angle of periapse g, respectively. The mean anomaly l has geometric significance; it is the ratio of the area swept out by the ray from the m_1-m_2 center of mass to the particle starting from its periapse to the total area in an inertial frame. The angle of periapse is also referred to as the argument of periapse. For more detail, see Szebehely [1967], Abraham and Marsden [2008], and Meyer and Hall [1992].

Delaunay Variables: Rotating Frame. As Figure 10.2.3 shows, $g = \bar{g} + t$ is the angle of periapse measured from the inertial axis x_{iner}, while $\bar{g}$ is the angle of periapse measured from the rotating axis x_{rot}. We prefer to use $\bar{g}$ instead of g as this gives rise to a time independent Hamiltonian function,

$$H = -\frac{1}{2L^2} - \mu R - G. \tag{10.2.4}$$

In what follows, we will describe the transformation from rotating cartesian coordinates and velocities $(x, y, \dot{x}, \dot{y})$ to the Delaunay canonical elements $(l, \bar{g}, L, G)$.

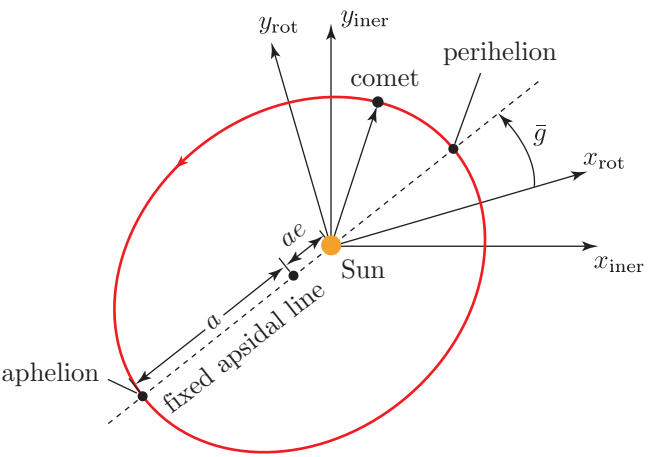


FIGURE 10.2.3. Geometry of the Delaunay variables. Elliptical orbits in the fixed (inertial) and rotating frames.

Transformation from Polar Coordinates to Delaunay Variables. We first transform from rotating cartesian coordinates and velocities $(x, y, \dot{x}, \dot{y})$ to rotating polar coordinates and velocities $(r, \theta, \dot{r}, \dot{\theta})$,

$$\begin{aligned} r &= \sqrt{x^2 + y^2}, \\ \theta &= \arctan\left(\frac{y}{x}\right), \\ \dot{r} &= \dot{x}\left(\frac{x}{r}\right) + \dot{y}\left(\frac{y}{r}\right), \\ \dot{\theta} &= -\dot{x}\left(\frac{y}{r^2}\right) + \dot{y}\left(\frac{x}{r^2}\right). \end{aligned} \tag{10.2.5}$$

The transformation $(r, \theta, \dot{r}, \dot{\theta}) \to (l, \bar{g}, L, G)$ is given by

$$\begin{aligned} G &= r^2(1 + \dot{\theta}), \\ L &= \frac{1}{\sqrt{-\left(\frac{G^2}{r^2}\right) + \left(\frac{2}{r}\right) - \dot{r}^2}}, \\ l &= \arccos\left(\frac{1 - \frac{r}{a}}{e}\right) - \left(\frac{r\dot{r}}{L}\right), \\ \bar{g} &= \theta - \arccos\left(\frac{\frac{G^2}{r} - 1}{e}\right) \end{aligned} \tag{10.2.6}$$

10.3 Lobe Dynamics and Resonant Gravity Assists

Lobe dynamics provides a general theoretical framework for discovering, describing and quantifying the transport "alleyways" connecting resonances (Dellnitz, Junge, Koon, Lekien, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005]). A resonance region and the lobes of phase space associated with movement around it are shown in Figure 10.3.1 on a Poincaré section in the same quasi-action-angle coordinates as Figures 10.2.1 and 10.2.2 and revealing the otherwise 'invisible' structure of the chaotic sea. The lobes are defined using the stable and unstable manifolds associated to unstable resonant orbits. Starting in one of the lobes above the resonance, an initial condition can get transported to below the resonance, and vice versa. This corresponds decrease or increase in the spacecraft's semimajor axis for zero fuel cost. Supposing our spacecraft is on the U_4 Poincaré section, lobe dynamics reveals those parts of the phase space which lead

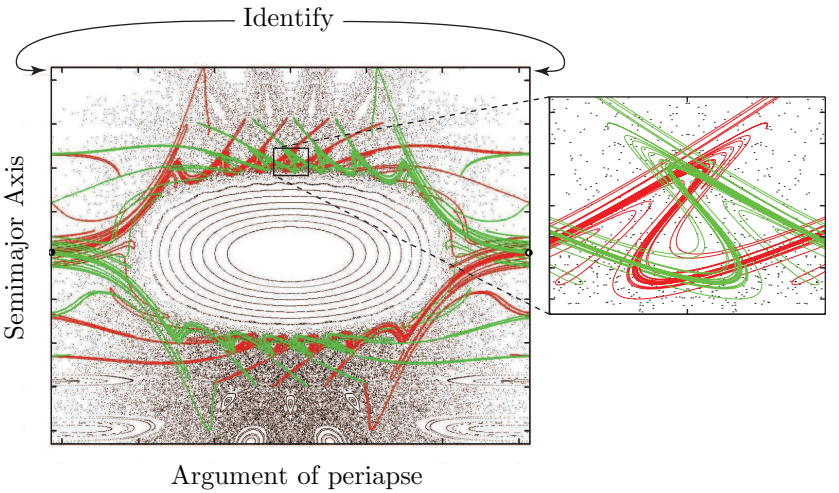


FIGURE 10.3.1. **Movement across an orbital resonance using lobe dynamics.** An unstable resonant orbit appears as a hyperbolic fixed point (the half-filled circles, identified as the same point) on this Poincaré section. The orbit's stable and unstable manifolds define both the resonance region (the central region with the target pattern) and "lobes" which transport phase points from above the resonance to below, and vice versa. An enlargement of the boxed region is shown at right.

to semimajor axis decreases or increases. As shown in Figure 10.3.2, those parts are sometimes close to each other.

In the language of astrodynamics, this corresponds to a gravity assists by the smaller primary. It is a special kind of gravity assist, however, because the resonant nature of the pre- and post-GA trajectory can lead to a series of such GAs, each one further decreasing or increasing the spacecraft's semimajor axis as the trajectory designer wishes.

Lobe Dynamics: Transport Between Regions. The determination of initial conditions which lead to desired GA sequences can be translated into the task of computing the *transport* between *regions* in phase space. More precisely, we consider a volume- and orientation-preserving map $f : \Sigma \to \Sigma$ (e.g., the Poincaré map in the PCR3BP as described in previous chapters) on some compact set $\Sigma \subset \mathbb{R}^2$ with volume–measure σ and ask for a suitable (i.e. depending on the application in mind) *partition* of Σ into compact *regions of interest* R_i, $i = 1, \ldots, N_R$, such that

$$\Sigma = \bigcup_{i=1}^{N_R} R_i \quad \text{and} \quad \sigma(R_i \cap R_j) = 0 \text{ for } i \neq j. \tag{10.3.1}$$

Following Rom-Kedar and Wiggins [1990], lobe dynamics theory states that the two-dimensional phase space Σ of the Poincaré map f can be divided as outlined above (see Eq. (10.3.1)), as illustrated in Figure 10.3.3(a).

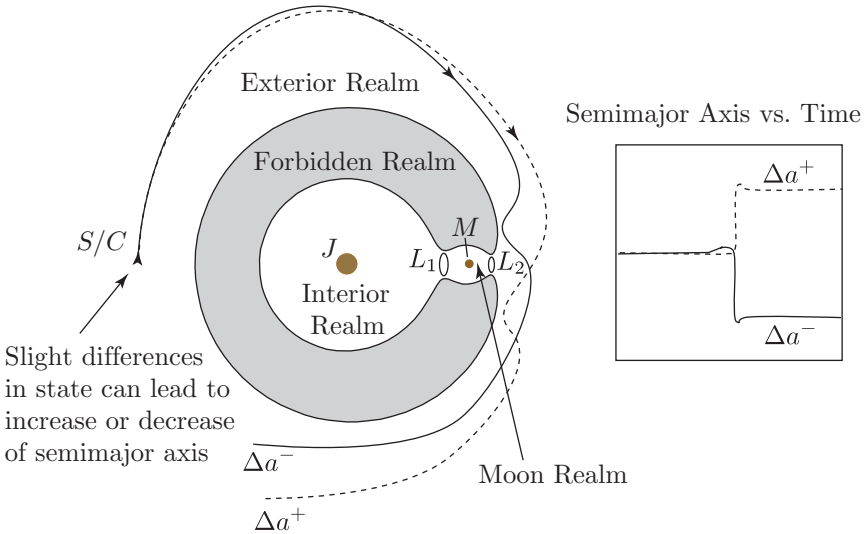


FIGURE 10.3.2. **Lobe dynamics reveals those parts of the phase space which lead to decreases or increases of semimajor axis.** We schematically show two initially neighboring spacecraft (S/C) trajectories which illustrate this effect in the presence of Jupiter (J) and one of its moons (M). The two trajectories, each with semimajor axis a^0, are initially on the U_4 Poincaré section and are in between two resonance regions. The solid trajectory is initially in a lobe which moves across the "lower" orbital resonance of semimajor axis $a^- < a^0$ and thus after the close approach to M the S/C has a lower semimajor axis. The change is approximately the width of the resonance region, Δa^-. The dotted trajectory is initially in a lobe which moves across the "upper" resonance and thus has a positive change of approximately the width of that resonance region, Δa^+. The energy regime shown is case 3, so there are bottlenecks around L_1 and L_2, through which orbits passing between realms must pass. But before passing between realms, the trajectory must cross a sequence of resonances first.

A region, R_j, is a connected subset of Σ with boundaries consisting of parts of the boundary of Σ (which may be at infinity) and/or segments of stable and unstable manifolds of hyperbolic fixed points, $p_i, i = 1, ..., N$. Moreover, the transport between regions of phase space can be completely described by the dynamical evolution of small regions of phase space, "lobes" enclosed by segments of the stable and unstable manifolds, as shown schematically in Figure 10.3.3(b), and defined below.

Boundaries, Regions, Pips, Lobes, and Turnstiles Defined. To define a *boundary* between regions, one first defines a *primary intersection point*, or *pip*. Looking at Figure 10.3.3(a), a point q_k is called a pip if $S[p_i, q_k]$ intersects $U[p_j, q_k]$ only at the point q_k, where $U[p_j, q_k]$ is a segment of the unstable manifold $W^u(p_j)$ joining the unstable fixed point p_j to q_k and similarly $S[p_i, q_k]$ is a segment of the stable manifold $W^s(p_i)$ of the unstable fixed point p_i joining p_i to q_k. The union of segments of the

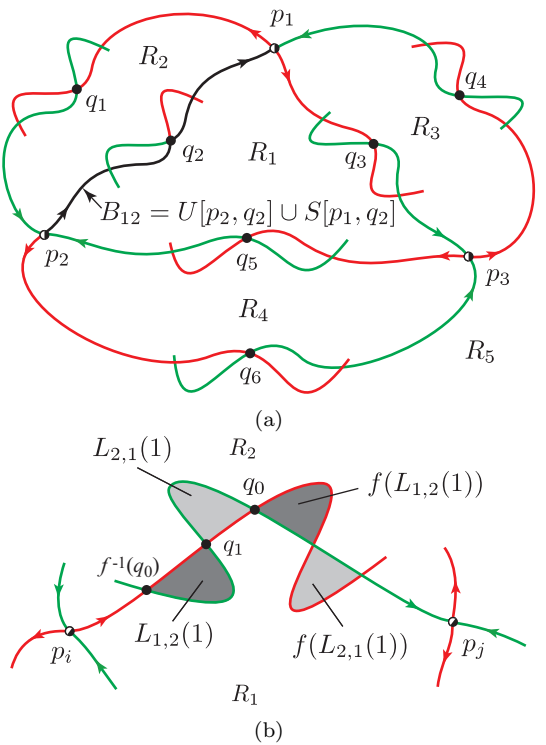


FIGURE 10.3.3. **Transport between regions of the phase space Σ of a Poincaré map f.** (a) The segment $S[p_1, q_2]$ of the stable manifold $W^s(p_1)$ from p_1 to q_2 and the segment $U[p_2, q_2]$ of the unstable manifold $W^u(p_2)$ from p_2 to q_2 intersect in the pip q_2. Therefore, the boundary B_{12} can be defined as $B_{12} = U[p_2, q_2] \bigcup S[p_1, q_2]$. The region on one side of the boundary may be labeled R_1 and the other side labeled R_2. (b) q_1 is the only pip between the two pips q_0 and $f^{-1}(q_0)$ in $W^u(p_i) \bigcap W^s(p_j)$, thus $S[f^{-1}(q_0), q_0] \bigcup U[f^{-1}(q_0), q_0]$ forms the boundary of precisely two lobes; one in R_1, labeled $L_{1,2}(1)$, and the other in R_2, labeled $L_{2,1}(1)$. Under one iteration of f, the only points that can move from R_1 into R_2 by crossing the boundary B are those in $L_{1,2}(1)$. Similarly, under one iteration of f the only points that can move from R_2 into R_1 by crossing B are those in $L_{2,1}(1)$.

unstable and stable manifolds naturally form partial barriers, or *boundaries* $U[p_j, q_k] \cup S[p_i, q_k]$, between the regions of interest $R_i, i = 1, ..., N_R$, in $\Sigma = \cup R_i$. In Figure 10.3.3(a) several pips are shown as well as the boundary B_{12}. Note that we could have $p_i = p_j$.

Consider Figure 10.3.3(b). Let $q_0, q_1 \in W^u(p_i) \cap W^s(p_j)$ be two adjacent pips, i.e., there are no other pips on $U[q_0, q_1]$ and $S[q_0, q_1]$, the segments of $W^u(p_i)$ and $W^s(p_j)$ connecting q_0 and q_1. We refer to the region interior to $U[q_0, q_1] \cup S[q_0, q_1]$ as a *lobe*. Then $S[f^{-1}(q_0), q_0] \cup U[f^{-1}(q_0), q_0]$ forms the boundary of precisely two lobes; one in R_1, defined by $L_{1,2}(1) \equiv \text{int}(U[q_0, q_1] \cup S[q_0, q_1])$, where int denotes the interior operation on sets,

and the other in R_2, $L_{2,1}(1) \equiv \text{int}(U[f^{-1}(q_0), q_1] \cup S[f^{-1}(q_0), q_1])$. Under one iteration of f, the only points that can move from R_1 into R_2 by crossing B_{12} are those in $L_{1,2}(1)$. Similarly, under one iteration of f the only points that can move from R_2 into R_1 by crossing B_{12} are those in $L_{2,1}(1)$. The two lobes $L_{1,2}(1)$ and $L_{2,1}(1)$ are together called a *turnstile*. It is important to note that $f^{-n}(L_{1,2}(1)), n \geq 2$, need not be contained entirely in R_1, i.e., the lobes can leave and re-enter regions with strong implications for the dynamics. The movement of points among regions, of which resonance regions are a kind, can be expressed compactly in terms of intersection areas of images or pre-images of turnstile lobes.

Transport Between Resonance Regions. In the application in the present chapter, the phase space Σ is known to possess resonance regions whose boundaries have lobe structures, which can lead to complicated transport properties (cf. MacKay, Meiss, and Percival [1984, 1987]; Meiss [1992]; Schroer and Ott [1997]; Koon, Lo, Marsden, and Ross [2000]).

The Poincaré section shown in Figure 10.2.1(b) of the resonance regions in the exterior realm is a good example. Suppose we consider a single resonance in the exterior realm, and we want to consider the phenomena of crossing a resonance and capture into the resonance using lobe dynamics. We can define the resonance region as outlined above, and then obtain a distribution of times for particles to cross from one side of the resonance to the other and residence times within the resonance region. The manifolds shown in Figure 10.3.1, as well as the associated pips, form the lobes that will determine the transport between three regions: R_2, the resonance region; R_1 and R_3, the regions below and above the resonance region, respectively. In Figure 10.3.4, the lobes for the upper and lower boundaries are shown. These boundaries each show a sequence of four lobes. Using the lobe dynamics framework, the movement of points between these regions can be computed systematically.

In fact, we can extend this methodology to two or more resonances. This is important for space missions, as fuel savings goes up significantly when multiple resonances are crossed using the natural dynamics. For example, consider the tour shown in Figure 1.2.10, particularly the outermost portion where Callisto is repeatedly encountered. The several Callisto flybys which constitute this portion of the tour exhibit roughly the same spacecraft-Callisto-Jupiter geometry. This portion of the trajectory takes several Callisto orbital periods. It culminates in a ballistic capture of the spacecraft by Callisto.

Merging the Theories of Lobe Dynamics and Tube Dynamics The work of Jaffé, Ross, Lo, Marsden, Farrelly, and Uzer [2002] is the first step toward an exact global theory of the transport. Such a picture requires the merging of tube dynamics with lobe dynamics: the transport between the interior, capture, and exterior realms is controlled by the tube dynamics; but the dynamics within the interior and exterior realms is dominated

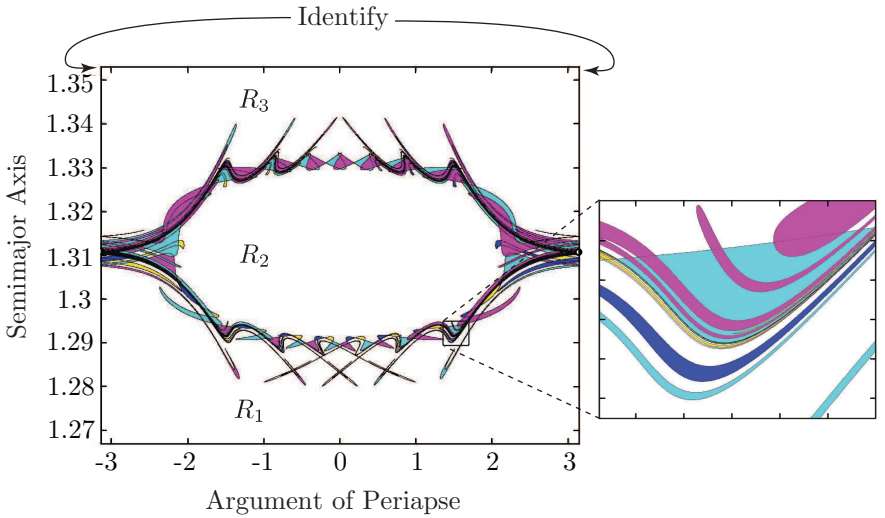


FIGURE 10.3.4. **Transport across a resonance.** The lobes formed by both the upper and lower boundaries of a resonance region R_2 are shown. Using the lobe dynamics framework, the transport of points between these regions can be computed systematically. An enlargement of the boxed region is shown at right.

by the lobe dynamics. Many interesting dynamical astronomy problems exhibit both kinds of dynamics, e.g., the exchange of impact ejecta between terrestrial planets (Gladman, Burns, Duncan, Lee, and Levison [1996]), and the evolution of scattered Kuiper belt objects into short period comets (Torbett and Smoluchowski [1990]; Malhotra, Duncan, and Levison [2000]).

For our purposes, we use tube dynamics along with lobe dynamics to find uncontrolled trajectories which quickly traverse the space between moons during the inter-moon transfer phase. Essentially, the lobes act as templates, guiding pieces of the tube across resonance regions. We can numerically determine the fastest trajectory from an initial region of phase space (e.g., orbits which have just escaped from moon M_2) to a target region (e.g., orbits which will soon be captured by a neighboring moon M_2). This yields the ΔV vs. TOF trade-off for the inter-moon transfer between Ganymede and Europa.

Remarks. Methods other than lobe dynamics which have been used by the dynamical astronomy community in studying problems related to transport between resonances can be broadly characterized as follows.

- *Large Scale Numerical Simulations.* In principle, the computation of rates of mass transport can be accomplished by numerical simulations in which the orbits of vast numbers of test particles are propagated in time including as many gravitational interactions as desirable. Many investigators have used this approach successfully (cf. Levison and

Duncan [1993]). However, such calculations are computationally demanding and it may be difficult to extract from them information about key dynamical mechanisms since the outcomes may depend sensitively on the initial conditions used for the simulation. In order to obtain *general features* of planetary system evolution and morphology, which is a major goal of dynamical astronomy, other approaches may be necessary.

- *Single Resonance Theory and Resonance Overlap Criterion.* One approach is to develop simple analytical models which provide answers to basic phase space transport questions. Much progress has been made in this area, but most of the work has focused on the study of the local dynamics around a single resonance, using a one-degree-of-freedom pendulum-like Hamiltonian with slowly varying parameters. Transport questions regarding capture into, and passage through resonance, have been addressed this way (see Henrard [1982], Neishtadt [1996], and Neishtadt, Sidorenko, and Treschev [1997])

  An important result regarding the interaction between resonances was obtained by Wisdom [1980]. In this paper, the method of Chirikov [1979] was applied to the PCR3BP in order to determine a resonance overlap criterion for the onset of chaotic behavior for small mass parameter, μ.

  These analytical methods are still used today (see Murray and Holman [2001] and references therein).

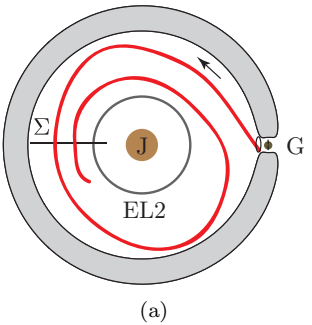
Using the nominal dynamics near a resonance in concert with efficient control to move phase points has been looked at in other applications as well, including the dynamics of charged particles in electromagnetic fields and mixing in fluids. See Vainchtein and Mezić [2004] for more information.

10.4 Trade-Off: ΔV vs. Time of Flight

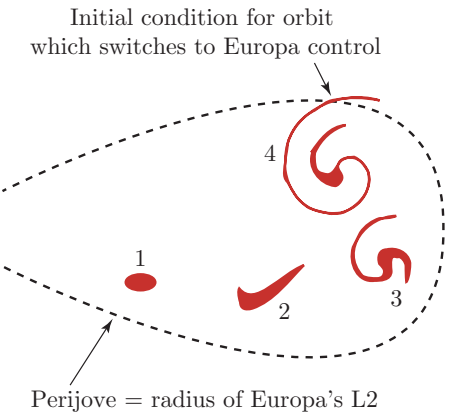
Method Description. Returning to the generation of inter-moon transfer, let us consider a transfer between orbits around Ganymede and Europa. We can initially consider an impulsive transfer from a Ganymede L_1 orbit (denoted Ga_{L1}) to a Europa L_2 orbit (Eu_{L2}). If we find such a transfer, we know that a transfer between orbits around Ganymede and Europa is nearby in phase space (Koon, Lo, Marsden, and Ross [2000]). We can consider the transfer in two pieces.

1. In the first piece, we consider the transfer along the unstable manifold tube of a Ga_{L1}, which we denote $U(\mathrm{Ga}_{L1})$. The object $U(\mathrm{Ga}_{L1})$ has

two branches, but we consider the one, as shown in Figure 10.4.1(a), heading initially in the direction of Europa's orbit.



(a)



(b)

FIGURE 10.4.1. **Numerical construction of natural trajectory arcs which will switch control from Ganymede to Europa.** Suppose we want to find trajectories which begin near Ganymede (G in the figure) and escape toward Europa, finally getting naturally captured by Europa. The first step is to numerically construct the Ganymede L_1 tube heading toward Europa, or $U(\mathrm{Ga}_{L1})$ in the terminology of the text. We take a Poincaré section, Σ, at the position shown in (a). We show only two crossings of Σ, but there are an infinite number. We also show the radial distance of Europa's L_2, labeled EL2, and the forbidden region at this energy. In (b), we show a schematic of the initial cross-section of the tube on Σ, labeled 1. The successive crossings are labeled 2, 3, ... In this schematic, we also show the dotted line corresponding to a perijove equal to the radial distance of Europa's L_2. The first three crossings are entirely within the zone of perijoves greater than the radial distance of Europa's L_2. Any spacecraft trajectory in the Jupiter-Ganymede-spacecraft system which crosses this line can be assumed to "switch" control to Europa, meaning the Jupiter-Europa-spacecraft system becomes a good approximation from then on. The coordinates represented here are not quasi-action-angle coordinates as in Figures 10.3.1 and 10.2.2(a), but are related to cartesian coordinates, which are easier to handle numerically.

The set of all Ga_{L1}'s is parameterized by the energy E_{Ga}, one of our tunable parameters. For each E_{Ga}, one can compute the Ga_{L1} and $U(\mathrm{Ga}_{L1})$. By taking a Poincaré section such as Σ in the figure, one can determine the trajectory within $U(\mathrm{Ga}_{L1})$ which takes the least time to transfer to a perijove distance r_p, equal to the approximate radial distance from Jupiter of Europa's L_2 point, labeled EL2 in the figure. The perijove distance line, the dotted line in the figure, can be computed as follows. Each point $(x, \dot{x}) \in \Sigma$ corresponds to an osculating conic orbit about Jupiter with an instantaneous semimajor axis a and eccentricity e. The dotted line in this figure is the locus of points satisfying $r_p = a(1 - e)$ for a perijove r_p equal to the approximate radial distance from Jupiter of Europa's L_2 point. Any point in Σ on this line is an orbit which can come close to Europa's L_2.

The computation of the trajectory which reaches this line in minimum time is performed numerically by determining the minimum tube crossing number on Σ which crosses the aforementioned perijove distance, the dotted line shown in Figure 10.4.1(b). Thus we find the natural trajectory arc which will switch "control" from Ganymede to Europa as the main perturber from jovicentric motion. The time of flight of this portion of the inter-moon transfer trajectory, T_{Ga}, is seen numerically to be a function of E_{Ga}.

2. For the second piece, we consider the transfer along the branch of the stable manifold tube of a Eu_{L2}, denoted $S(\mathrm{Eu}_{L2})$, heading initially in the direction of Ganymede's orbit. The set of all Eu_{L2}'s is parameterized by the energy E_{Eu}, another tunable parameter. For each E_{Eu}, one can compute the Eu_{L2} and $S(\mathrm{Eu}_{L2})$. One can determine the trajectory within $S(\mathrm{Eu}_{L2})$ which takes the least time to transfer to an apojove distance r_a, equal to the approximate radial distance from Jupiter of Ganymede's L_1 point. The time of flight of this trajectory, T_{Eu}, is found numerically to be a function of E_{Eu}.

The sum, $\mathrm{TOF} = T_{\mathrm{Ga}} + T_{\mathrm{Eu}}$, is an approximate inter-moon transfer time. The total fuel expenditure, ΔV_{tot}, needed to perform the transfer can be estimated as follows. We assume only two impulsive maneuvers, ΔV_{Ga} and ΔV_{Eu}.

$\Delta V_{\mathrm{Ga}} =$ the ΔV to escape from the scientific orbit around Ganymede, which can be estimated using the energies of the transfer away from Ganymede, E_{Ga}, and of the scientific orbit at Ganymede, E_{GaO}.

$\Delta V_{\mathrm{Eu}} =$ the ΔV to enter the scientific orbit around Europa, which can be estimated using energies of the transfer toward Europa, E_{Eu}, and of the scientific orbit at Europa, $E_{\mathrm{Eu}}O$.

The total fuel expenditure is the sum, $\Delta V_{\mathrm{tot}} = \Delta V_{\mathrm{Ga}} + \Delta V_{\mathrm{Eu}}$. We suppose that E_{GaO} and E_{EuO} are given. We can then perform this procedure

for a range of tunable parameters (E_{Ga} and E_{Eu}), to determine the fuel consumption (ΔV_{tot}) versus time of flight (TOF) trade-off.

Computing the ΔV's. We assume that portions of each tube quickly reach a periapse of 100 km altitude above each moon, and that the solutions which do this are close in phase space to the transfer solutions found, assumptions justified by earlier work (Koon, Lo, Marsden, and Ross [2000]; Gómez, Koon, Lo, Marsden, Masdemont, and Ross [2001]). Given these assumptions, we can estimate ΔV_{Ga} and ΔV_{Eu} as follows. In the rotating frame of a Jupiter-moon-spacecraft three-body system, a spacecraft with a velocity magnitude v has a three-body energy

$$E = \frac{1}{2}v^2 + \bar{U}, \tag{10.4.1}$$

where the effective potential, a function of position, is

$$\bar{U} = -\frac{1}{2}r^2 - \frac{1-\mu}{r_J} - \frac{\mu}{r_M} - \frac{1}{2}\mu(1-\mu), \tag{10.4.2}$$

where μ is the mass ratio $\frac{m_M}{m_J+m_M}$, r_J is the spacecraft's distance from Jupiter's center, r_M the spacecraft's distance from the moon's center, and r the spacecraft's distance from the Jupiter-moon center of mass, which is very close to Jupiter. At a distance of 100 km altitude above the moon, we are very close to the moon. Therefore, using the standard non-dimensional units, $r \approx r_J \approx 1$, and we can approximate (10.4.2) as

$$\begin{aligned}\bar{U} &\approx -\frac{1}{2}(1)^2 - \frac{1-\mu}{1} - \frac{\mu}{r_M},\\ &\approx -\frac{1}{2} - 1 - \frac{\mu}{r_M},\\ &\approx -\frac{3}{2} - \frac{\mu}{r_M}.\end{aligned} \tag{10.4.3}$$

Using (10.4.1), the velocity can then be approximated as

$$v \approx \sqrt{2\left(\frac{\mu}{r_M} + \frac{3}{2} + E\right)}. \tag{10.4.4}$$

Therefore, the approximate ΔV to go between energies E_a and E_b while at the same distance $r_M \ll 1$

$$\Delta V \approx \left|\sqrt{2\left(\frac{\mu}{r_M} + \frac{3}{2} + E_a\right)} - \sqrt{2\left(\frac{\mu}{r_M} + \frac{3}{2} + E_b\right)}\right|. \tag{10.4.5}$$

We can use the above equation to compute ΔV_{Ga} given $E_a = E_{\mathrm{Ga}}$ and $E_b = E_{\mathrm{GaO}}$. For this study, we take E_{GaO} to be the energy of L_1 in the

Jupiter-Ganymede-spacecraft system. This corresponds to a bound elliptical orbit around Ganymede which is at the energy threshold of escape, and therefore cannot escape the Hill region around Ganymede. We can perform similar calculations for ΔV_{Eu}.

The result of tabulating $\Delta V_{\mathrm{tot}} = \Delta V_{\mathrm{Ga}} + \Delta V_{\mathrm{Eu}}$ for each TOF $= T_{\mathrm{Ga}} + T_{\mathrm{Eu}}$ is given in Figure 10.4.2(a).

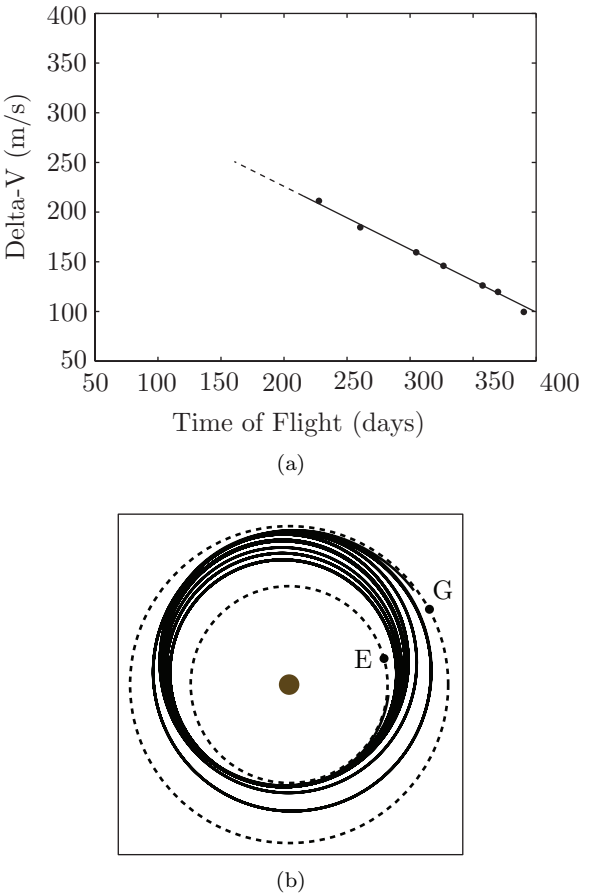


FIGURE 10.4.2. **Fuel consumption versus flight time trade-off for the inter–moon transfer phase of a multi-moon orbiter mission.** (a) The ΔV vs. time of flight plot for several transfer trajectories from Ganymede to Europa are shown. For the several cases run, we find a near linear relationship between ΔV and time of flight. For this study we looked at a range of energies in both three-body systems. The highest energy (and lowest TOF) transfer we computed is shown in (b) in inertial coordinates, where G labels Ganymede's orbit and E labels Europa's. This transfer has a TOF of 227 days and a ΔV of 211 m/s. Beyond this lower TOF limit to the computations, we speculate that the linearity will continue for a while, indicated by the dashed line.

In dimensional units, we find a near linear relationship between ΔV_{tot} and TOF, given roughly by

$$\Delta V_{\text{tot}} = 340 - 0.60 \times \text{TOF}, \tag{10.4.6}$$

where ΔV_{tot} is given in m/s and TOF is given in days.

For this study we looked at a range of energies in both three-body systems. The highest energy (and lowest TOF) transfer we computed is shown in Figure 10.4.2(b) in inertial coordinates, where G labels Ganymede's orbit and E labels Europa's. This transfer has a TOF of 227 days and a ΔV of 211 m/s. Beyond this lower TOF limit to our computations, we speculate that the linearity will continue for a while, indicated by the dashed line. Further computations are needed to settle this matter.

Transfers between Low Altitude, Circular Orbits. In the above, we have computed only the minimum ΔV necessary to go between bound orbits around each moon. It is instructive to note the additional ΔV which would be necessary to effect a transfer between low altitude circular orbits of zero inclination. According to Villac and Scheeres [2001] who used the Hill three-body problem for their model, the minimal ΔV to escape a low altitude (around 100 km) Ganymede would be at most 669.5 m/s, and to inject into a low altitude (around 100 km) Europa orbit would be at most 451.2 m/s. Therefore, an additional $\Delta V_{\text{circ}} = 1120.7$ m/s can be added to each point in the curve of Figure 10.4.2(a) in order to approximate a transfer between zero inclination, low altitude circular orbits. For example, the transfer with a 227 day TOF requires a total ΔV of about 1332 m/s to transfer between circular orbits.

Some Remarks. As demonstrated, minimum time, two-impulse transfers from Ganymede to Europa in the framework of the patched three-body approximation can be investigated as a function of the three-body energy. The transfers are between orbits bound each moon, respectively. Tube dynamics have been used along with lobe dynamics to find uncontrolled (natural) trajectories which quickly traverse the space between the moons. The lobes act as templates, guiding pieces of the tube across resonance regions. The tubes have the numerically observed property that the larger the energy, the further the tube travels from its associated libration point in a fixed amount of time. This property has been exploited to find the time of flight between Ganymede and Europa as a function of the energy in their respective three-body systems. The energies have been used to calculate the ΔV of escape, or get captured, from each moon, respectively. Our results show that in the range of energies studied, the ΔV vs. time of flight relationship is nearly linear and that a reasonable inter-moon time of flight for a multi-moon orbiter can be achieved using a feasible ΔV. The determination of this relationship for a larger range of energies would be helpful. As discussed below, a linear relationship between ΔV and flight time is by no means generic for low energy trajectories.

The ΔV vs. TOF Curve for an Earth-to-Moon Trajectory. The importance of the ΔV vs. TOF curve is not limited to MMOs. A case of a non-linear trend is shown below for a set of Earth-to-Moon trajectories. Consider a family of trajectories from an fixed Earth orbit to a fixed lunar orbit. Bollt and Meiss [1995] considered the transfer from a circular Earth orbit of radius 59669 km to a quasi-periodically precessing ellipse around the moon, with a perilune of 13970 km. Their method takes advantage of the fact that long trajectories in a compact phase space are recurrent. Starting with a long unperturbed chaotic trajectory that eventually reaches the target, the use small well chosen ΔV's to cur recurrent loops from the trajectory, shortening it whenever possible. They find a transfer (see Figure 10.4.3(a))

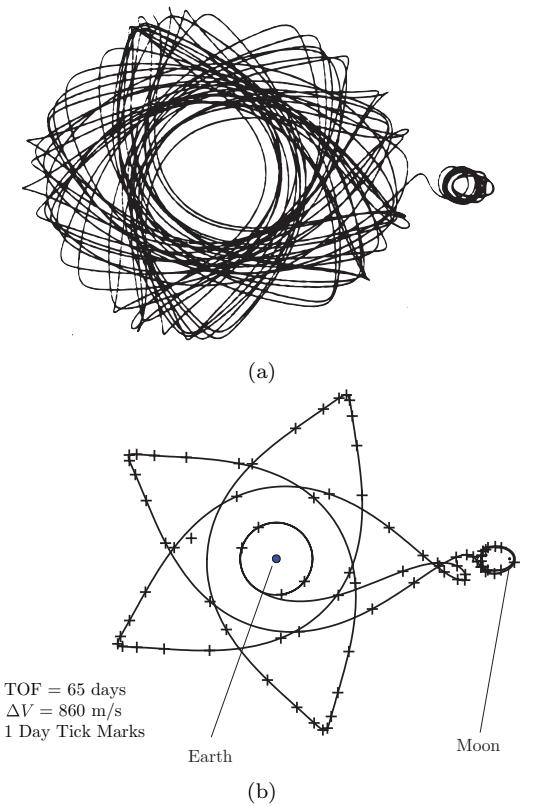


FIGURE 10.4.3. **Earth-to-Moon trajectory: much lower TOF achieved using only slightly more fuel.** (a) The transfer from a circular earth orbit of radius 59669 km to precessing lunar orbit of perilune 13970 km found by Bollt and Meiss [1995] is shown in the rotating frame. The ΔV is 749.6 m/s and the TOF is 748 days. (b) A transfer between the same initial and final orbits, using a ΔV of 860.1 m/s, but requiring a TOF of only 65 days.

that achieves ballistic capture requiring 749.6 m/s, 38% less total velocity boost than a comparable Hohmann transfer, but requiring a transfer time of 748 days. Later, Schroer and Ott [1997] considered this problem with the same initial and final orbits, but found a transfer requiring about half the flight time, 377.5 days, but using roughly the same total ΔV, 748.9 m/s, suggesting that this is near the minimum required for a transfer between these two orbits.

Using the methods outlined earlier in this section, we can construct a family of transfer trajectories that provide a compromise between time and fuel optimization. We find a transfer, shown in Figure 10.4.3(b), with a flight time of 65 days which uses a total ΔV of 860.1 m/s. Thus the trajectory takes one-tenth of the time as the trajectory found in Bollt and Meiss [1995] and uses only about 100 m/s more fuel. The previously described trajectories are shown in Figure 10.4.4 compared with the family of trajectories produced using the method outlined here. The lower bound in ΔV, near 750 m/s, comes from the three-body energy difference between the initial and final orbits. Given a boost of that minimum amount while on the initial orbit, the resulting higher energy trajectory will wander through phase space, coming close to the final lunar orbit if one waits a sufficiently long amount of time. By using the phase space structures as done here, one can design trajectories that circumvent the extraneous wandering.

10.5 Some Open Problems for Multi-Moon Orbiters

A partial list of open areas of research regarding MMOs are the following.

- *Transferring between large inclination orbits about moons:* Future studies addressing the important compromise between time and fuel costs for a MMO mission will need to address the three-dimensionality of the problem. A likely requirement for a MMO may be the necessity to go from an inclined orbit about one moon to an inclined orbit about another, such as shown in Figure 1.2.10(b). We speculate that this may lead a nearby curve, possibly linear, in the ΔV vs. TOF plot.

- *The use of low thrust continuous propulsion and optimal control:* Low thrust trajectory control, described in the next section, is of great interest to current mission design. The current work on MMOs as discussed above considers several small impulsive burns. But an actual mission may want to save on spacecraft weight by using low thrust propulsion. How could the methods outlined here be modified to incorporate low thrust? Theoretically, one of the most favored approaches is to use optimal control in generating low thrust trajecto-

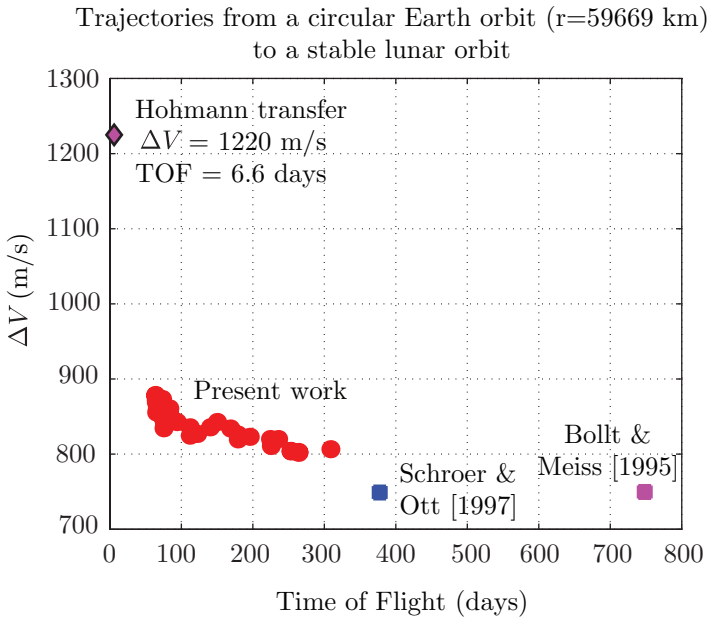


FIGURE 10.4.4. **Trade-off between fuel and time optimization for an Earth–to-Moon trajectory.** The ΔV vs. time of flight plot for several transfer trajectories to the moon which use nonlinear third-body effects, compared with the Hohmann transfer. A trajectory of one-fifth to one-tenth of the flight-time of some previous fuel optimized trajectories can be achieved using only about 100 m/s more ΔV.

ries. We have found that a good first guess is often vital for numerical optimization algorithms, especially for an N-body problem, which is numerically very sensitive. Dynamical systems theory can provide geometrical insight into the structure of the problem and even good approximate solutions, as found in Serban, Koon, Lo, Marsden, Petzold, Ross, and Wilson [2002]. There is evidence that optimal trajectories using multiple low thrust burns are "geometrically similar" to impulsive solutions (Schoenmaekers, Horas, and Pulido [2001]; Zondervan, Wood, and Caughy [1984]). Thus, multiple burn impulsive trajectories that we construct for the MMO can be good first guesses for an optimization scheme which uses low thrust propulsion to produce a fuel efficient mission.

- *Radiation effects:* The current model does not include radiation effects. It is desirable to keep the spacecraft outside of a 12 R_J from Jupiter, in which the radiation may destroy sensitive electronics on board the spacecraft. The orbit of Europa is located at 10 R_J, so the transfer between Ganymede and Europa must minimize the time spent near its perijove for the final resonant gravity assists that lead

to a capture by Europa. One needs to determine what is the best way to minimize radiation effects and still achieve a low energy, low thrust transfer. On the other hand, the strong magnetic field of Jupiter may make the use of tethers a viable propulsion or power generation option.

- *More control over operational orbits for scientific observation:* For a mission to Europa and the other moons, some control strategy is necessary to maximize desirable scientific observation and avoid collisions with the moon surface or escape from the moon's vicinity. Exotic strategies might be considered. For instance, what is the optimal thrusting strategy during the ballistic capture approach in order to achieve an operational orbit which maximizes observation time over an interesting portion of a moon's surface? Also, what is the optimal station-keeping strategy for elliptical operational orbits? In the short term, it may be desirable to target particular stable operational orbits, but still save fuel using third body effects (Gómez, Koon, Lo, Marsden, Masdemont, and Ross [2001]; Scheeres, Guman, and Villac [2001]).

- *Autonomous on-board navigation and control:* A trajectory using ballistic capture/escape and resonant GAs is sensitive to ΔV errors and modeling errors. A spacecraft executing such a trajectory will need to have the capability of autonomous on-board navigation and control. The first step toward this which one can look at is the trajectory correction maneuver problem as considered in Chapter 8, in which errors are modeled and a control algorithm corrects for those errors.

10.6 Low Thrust Trajectories

The maturity of current ion engine technology has brought low thrust controls into the practical world of mission design in industry and in NASA (cf. the Deep Space 1 mission). Similar work is being done at the European Space Agency (cf. the SMART-1 lunar mission).

The design of spacecraft trajectories characterized by continuous thrust levels that are very low compared to the vehicle weight represent a challenging class of optimization problems. *Low thrust trajectories* are demanding because realistic forces due to oblateness, and third-body perturbations often dominate the thrust. Furthermore, because the thrust level is so low, significant changes to the orbits require very long duration trajectories. Using the small continuous control thrust efficiently will have important implications for multi-moon orbiter and other missions in the future which intend to use low thrust technology.

The low thrust regime is defined by the maximum acceleration f_{max} of the spacecraft, the ratio of the maximum thrust (F_{max}) to mass (m),

$$f_{max} = \frac{F_{max}}{m}. \tag{10.6.1}$$

If this ratio is at or below approximately $10^{-2}g$, where $g = 9.7803267$ m/s^2 is the acceleration due to gravity on the Earth's equator, then the thrust level is deemed "low thrust" (Chobotov [1996]).

Low thrust trajectories, unlike the high thrust trajectories we have discussed previously, are trajectories which, while the thrust is on, continuously change the spacecraft's three-body energy, eq. (2.3.13). One thus can imagine a picture where the Hill's region changes in time. The most efficient low thrust trajectories will be those which go from, say, the exterior realm to a close orbit around a moon in the capture realm using the least fuel, as in Figure 10.6.1.

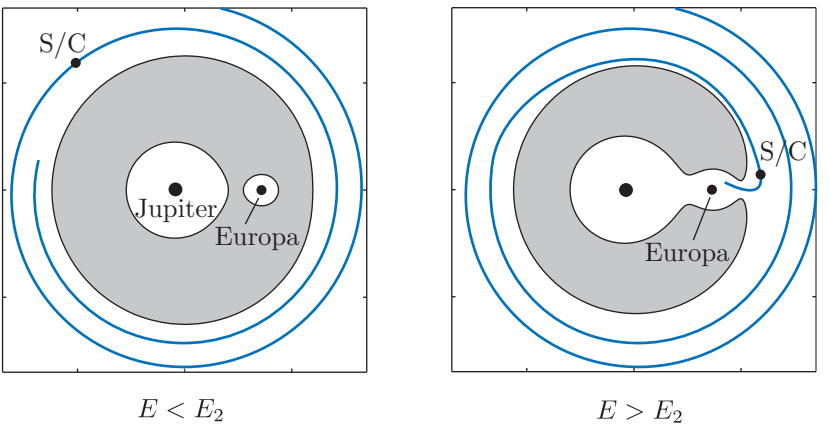


FIGURE 10.6.1. **Changing three-body energy via low thrust control.** Suppose that we begin in the exterior region of the Jupiter-Europa-spacecraft three-body system. If our goal is to spiral in on a Jupiter-centered orbit until a capture at Europa is effected, we turn on the thrust with the thrust vector anti-parallel to the velocity. We monitor the three-body energy E of the spacecraft (S/C), which is changing continuously as we thrust. This energy, which is constant for the motion of the S/C in the field of Jupiter and Europa in the case of no-thrust (coasting), can easily be computed via eq. (2.3.13). For most of the spiral-in, the L_2 neck through which the S/C will reach Europa is closed (left figure). The neck opens up above the energy of L_2, E_2 (right figure). We illustrate this in the Jupiter-Europa rotating frame.

In other words, one wants to use low thrust so that a spacecraft will go through the neck when it is narrowest. It makes sense to begin an extension of the dynamical systems methods we have discussed thus far to this situation where the energy, and thus the tube and lobes, are now changing.

Naive Approach: Heliocentric to Callisto Example. Based on this simplified picture of continuously changing energy surfaces, one can construct low thrust spacecraft (S/C) trajectories using control laws which are simple to implement. Consider the problem of obtaining a scientific orbit about Callisto from a near-Earth but heliocentric orbit. One would start with an arrival time and desired S/C orbit about Callisto, obtained from mission requirements. One numerically integrates in reverse time applying the maximum thrust, i.e., $||\mathbf{F}|| = F_{max}$, anti-parallel to the orbit velocity vector $\dot{\mathbf{x}}$. In reverse time the orbit will "spiral out," increasing the semi-major axis, three-body energy, and apoapse (see Kluever [1998]). The desired orbit about Callisto has no constraints, i.e., the method is valid for all desired orbits, such as polar orbits which are useful for scientific investigation of a body. As the three-body energy increases (in reverse time) and the orbit apoapse nears the Hill radius, the necks about L_1 and L_2 open. When these necks open, reverse time escape from Callisto to an orbit about Jupiter is permitted, as shown in Figure 10.6.2(a). We will refer to this orbit about Jupiter as the *pre-capture orbit*, since it leads to a capture into an orbit about Callisto in forward time.

If a pre-capture orbit about Jupiter with a Jupiter-centered semi-major axis larger than Callisto's semi-major axis is desired, then the reverse time trajectory should proceed through the L_2 neck, which is the case for Figure 10.6.2(a), with the pre-capture orbit shown in Figure 10.6.2(b). If a pre-capture orbit with a semi-major axis less than Callisto's is desired, then the reverse time trajectory should proceed through the L_1 neck. An L_1 or L_2 neck escape can be obtained by finely tuning the duration of the thrusting portion during the Callisto-centered reverse time escape before a coast period begins (in which no thrust is applied, $\mathbf{F} = 0$) in reverse time. The sensitivity of the orbital dynamics of the S/C when it reaches the Hill radius makes escape toward *either* the L_1 or L_2 neck possible. The appropriate thrust duration can be found by a simple scanning procedure in which one adjusts the duration and records whether an L_1 or L_2 neck escape occurred.

By integrating backward from a Callisto orbit and forward from an Earth orbit to a common intersection state, we are able to construct an end-to-end low-thrust Earth to Callisto trajectory. Such a phase would be necessary for any trajectory to sequentially orbit Jupiter's icy moons. By adjusting a small number of parameters using an optimizer, one is able to reach a common intersection state.

This naive approach can also be used for computing the transition between moons. For example, one computes a Ganymede capture trajectory in much the same way as the escape trajectory from Callisto (except for time reversal). This would guarantee a capture onto a desirable Ganymede scientific orbit from a desirable Callisto scientific orbit. When the capture trajectory (going backward in time) reaches a Jupiter-centered orbit heading toward Callisto in a Jupiter-centered inertial reference frame, we label

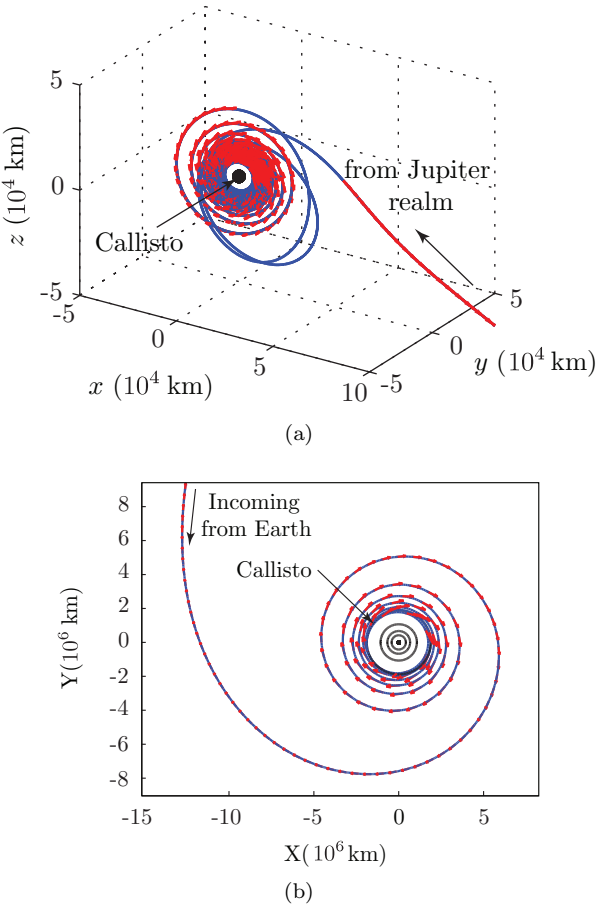


FIGURE 10.6.2. **A low-thrust trajectory from Earth to a Callisto polar orbit** obtained using the naive approach. (a) The spacecraft (S/C) trajectory is shown in a Callisto-centered inertial frame. The arrows along the S/C trajectory indicate the thrust direction. Coasting arcs are given by the absence of arrows. The outer most portion of the Callisto capture spiral is shown. The trajectory will spiral into a desired orbit at the desired arrival time. The S/C capture trajectory was obtained by reverse time integration from the desired orbit at the desired epoch. In reverse time, the S/C "escapes" to an orbit around Jupiter in the Jupiter realm through the L_2 neck. In forward time, it is captured by Callisto from a pre-capture orbit around Jupiter, which is shown in (b). (b) The same S/C trajectory is shown in a Jupiter-centered inertial frame, and extending further back in time to where the S/C entered the Jupiter realm from the Sun realm. This pre-capture orbit originated at Earth and has a very low inclination with respect to Jupiter. The arrows along the S/C trajectory indicate the thrust direction. The orbits of the four Galilean moons of Jupiter are shown. The spiral is approaching Callisto's orbit and will enter the realm around Callisto through the L_2 neck.

this time and state $(\mathbf{x}_T, \mathbf{v}_T, t_T)$, where the subscript "$T$" stands for target. For the Jupiter-centered transition portion between the two moons, we would use a control law to target $(\mathbf{x}_T, \mathbf{v}_T, t_T)$, as in a rendezvous problem. The motion of the S/C in the Jupiter-centered transition portion is essentially a Kepler problem (the S/C in the field of Jupiter) with the moons as minor perturbations. There are known techniques for solving the rendezvous problem, as touched on in §8.3 and these could be implemented.

Furthermore, solving the rendezvous problem would also shed light on the trajectory correction problem to correct for control or modeling errors (e.g., re-design of nominal trajectory portions vs. returning to nominal trajectory portions).

A natural next step is to consider guidance, navigation, and control issues for a candidate low thrust MMO trajectory such as the one presented here. Toward this end, one can follow the methodology presented in Desai, Bhaskaran, Bollman, Halsell, Riedel, and Synnott [1997] and incorporate the appropriate algorithms into an overall software package.

Other Approaches, Open Problems, and Remarks.

- *Lyapunov-based transfer between orbits in the restricted three-body problem.* The naive approach can be improved upon if more information regarding the phase space structure is incorporated into the control law design. One area of possible study is to extend the two-body work of Chang, Chichka, and Marsden [2002] to a three-body context and consider the transfer of satellites between orbits of the restricted three-body problem using Lyapunov stability theory specific to this problem. The goal is to find a Lyapunov function, which gives a feedback controller such that the target orbit becomes a locally asymptotically stable periodic orbit in the closed-loop dynamics. One could start first in the Hill problem for simplicity.

- *Systematic study of transfers related to distant retrograde orbits.* Distant retrograde orbits (DROs) are three-body orbits that encompass the secondary body and lie entirely outside the libration points L_1 and L_2. Whiffen [2003] studied the process of alternate capture and escape trajectories in the context of a low thrust MMO of the Galilean moons and identified the dynamics near unstable DROs as particularly suitable for capture and escape. He found this to be especially true if a low retrograde orbit about a moon is an ultimate target. While Whiffen's observations are intriguing, a systematic study of the dynamics associated with DROs to further exploit this phenomenon has yet to be performed, e.g., their invariant manifolds and how they connect with realms around the moons and resonances between moons. They may have an important role similar to the role of periodic orbits and NHIMs related to libration points which we have discussed in previous chapters.

- *Trajectory correction maneuvers for low thrust trajectories in a three-body context.* As low thrust trajectories in a multi-body environment can be unstable to perturbations, one needs to consider model uncertainty, unmodeled dynamics, and noise. Moreover, one wants to answer the following question: when errors occur, should one re-target the original (nominal) trajectory or re-generate a new "nominal" trajectory?

- *Merging optimal control with dynamical systems theory.* The construction of exotic spacecraft orbits using tube and lobe dynamics requires optimal thruster controls to navigate dynamically sensitive regions of phase space. Using optimal, fuel minimizing impulsive and continuous thrust, is the most efficient and natural way to take advantage of the delicate dynamics.

  Lawden [1963] developed Primer Vector Theory, the first successful application of optimal control theory to minimize fuel consumption for trajectories with impulsive thrusts in the two-body problem. The extension of Primer Vector Theory to continuous low-thrust control for the restricted three-body problem is a current area of active research. Our work on this problem indicates that developing optimal control theory within the dynamical systems framework shows promise for producing a numerical solution in the three-body context.

  In our ongoing effort to use the methods of optimal controls to study the orbit transfer problem for certain JPL space missions, we are exploring the "direct" method for solving the optimal control problem. In the direct method, the optimal control problem can be first approximated by a discrete optimization problem using a collocation or multiple shooting discretization scheme. Then the resulting optimization problem is solved numerically with a sophisticated sequential quadratic programming (SQP) technique. While the numerical algorithm of the direct method is quite robust for certain types of two-body problems, we do not expect that application to the three-body regime will be completely straightforward. It would also be interesting to explore the ways in which optimal control in the presence of mechanics (as in, for example, Koon and Marsden [2000]) is useful in this problem.

  As usual, for any numerical algorithm, a good initial guess is vital, especially if the problem is very sensitive numerically. Dynamical systems theory can provide geometrical insight into the structure of the problem and even good approximate solutions. Clearly, this theoretical insight and its derivative numerical tools can aid in the construction of superior initial guesses that lead to a convergent solution.

  A deeper understanding of the dynamical structure of the restricted three-body problem, including the ideas we have contributed in this

book, may suggest alternative formulations of the optimizing scheme which are based more on the geometry of the phase space. Instead of "numerically groping in the dark," algorithms could be developed with the natural dynamics built in, thereby yielding better convergence properties.

- *Discrete mechanics and optimal control.* A new approach to the solution of optimal control problems for aerospace and mechanical systems shows great promise for low thrust missions. Known as *discrete mechanics and optimal control* (DMOC), it is based on a direct discretization of the Lagrange-d'Alembert principle of the system instead of the associated forced Euler–Lagrange equations. The resulting forced discrete Euler–Lagrange equations then serve as constraints for the optimization of a given cost functional.

  Several benefits of using this approach are expected. For instance, in variational integrators, one respects the energy of forced and damped systems much better than with standard algorithms. Correspondingly, the energy budget of a control system is expected to be more accurately computed using DMOC, and this will be especially so for long duration simulations, such as long low-thrust missions. In addition, one expects DMOC methods to be more robust to modeling errors.

Appendix A
Related Literature

A.1 How This Book Fits into the Larger Body of Literature

The work covered in this book is connected to a larger picture which we take a few pages to cover here to direct the interested reader to the appropriate references for additional information.

Of course there are many specific references to journal articles on dynamical systems theory and mission design throughout the book. Of works in book form, we especially direct the reader to the four volume set *Dynamics and Mission Design Near Libration Points* by Gómez, Jorba, Llibre, Martínez, Masdemont, and Simó [2001] and *Capture Dynamics and Chaotic Motions in Celestial Mechanics* by Belbruno [2004] which both complement the present work.

In the remainder of this section, the focus is on topics and relevant literature whose scope just touches the subject presented in this book.

Multiscale Dynamics and Phase Space Transport. The subject of dynamical systems has experienced considerable theoretical growth as well as having made significant applications, mainly motivated by recent progress in the development of numerical techniques for dynamical problems and the availability of more powerful computational facilities. However, many interesting and challenging problems remain. One area of particular interest is *phase space transport*, a unified mathematical description of dynamical processes which can be applied to a wide range of physical

phenomena across many scales, such as atomic physics (Jaffé, Farrelly, and Uzer [1999]), physical chemistry (Davis and Gray [1986]; De Leon and Ling [1994]; Martens, Davis, and Ezra [1987]; Vela-Arevalo [2002]; Burbanks, Waalkens, and Wiggins [2004]; Gabern, Koon, Marsden, and Ross [2005]), fluid mixing (Haller and Mezic [1998]; Poje and Haller [1999]; Haller and Yuan [2000]; Haller [2001]; Lekien [2003]), climate models (Pierrehumbert [1991a,b]), asteroid and comet evolution (Koon, Lo, Marsden, and Ross [2000, 2001a]; Ross [2003, 2004]; Koon, Marsden, Ross, Lo, and Scheeres [2004]; Dellnitz, Junge, Koon, Lekien, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005]; Dellnitz, Junge, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005]), stellar motion (Hénon and Heiles [1964]; Fordy [1991]), and cosmological models of large scale mass distribution in the universe (De Oliveira and Damião Soares [1998]; De Oliveira, Sauto, Damião Soares, and Tonini [2001]).

These physical systems can be modeled initially as non-integrable dynamical systems, and many are Hamiltonian with $n \geq 2$ degrees of freedom (n dof). Simulation and theoretical understanding of this rich class of problems is important to many areas of science and engineering. The restricted three-body problem is just one problem in this large class of systems which have a global *mixed phase space* structure of stable and chaotic zones: KAM tori and chaotic sea (cf. MacKay, Meiss, and Percival [1984, 1987]; Meiss and Ott [1986]; Malhotra [1996]; Malhotra, Duncan, and Levison [2000]).

To realistically model some problems requires $n \geq 3$ dof (Koon, Marsden, Ross, Lo, and Scheeres [2004]; Wiesenfeld, Faure, and Johann [2003]), i.e., a phase space of six dimensions or more. This high dimensionality has made the systematic study of such systems difficult.

However, recent work by the authors and others on Hamiltonian systems of 3 dof (Gómez, Koon, Lo, Marsden, Masdemont, and Ross [2004]; Gabern, Koon, Marsden, and Ross [2005]) has led to some progress using a variety of important semi-analytical and numerical tools. Moreover, there are some indications that this class of problems contains the essential elements of n dof, $n \geq 4$, systems. For example, in 3 dof the KAM tori no longer have the dimensionality to bound volumes of phase space, and evidence suggests the co-dimension one manifolds associated to bound invariant objects provide the important barriers to transport—partial barriers. There are at least two fundamental kinds of partial barriers, associated with two spatio-temporal dynamical scales, related to tube and lobe dynamics, respectively that we discuss in what follows.

This book fits in as one component of an overall research effort to develop new analytical and computational tools to treat phase space transport qualitatively and quantitatively, for Hamiltonian systems with three or more degrees of freedom. Some of the important techniques are as follows.

Invariant Manifolds, Tube and Lobe Dynamics. For many systems, the Hamiltonian is of the form kinetic plus effective potential. The shape of the potential provides information about unstable and stable equilibria as well as the geometry of realms surrounding equilibria, which partition the energy manifold on the coarsest level. The energy value determines the connectivity of the realms. For a rank one saddle, the transport between realms is mediated by *phase space tubes* bounded by the stable and unstable manifolds associated to normally hyperbolic invariant manifolds around unstable equilibria. The theory of *tube dynamics* developed in this book (building from Koon, Lo, Marsden, and Ross [2000] and numerous later papers) can be used as a basis of a statistical theory for the computation of transport rates between realms. An illustration of tube dynamics for a three degree of freedom system is shown in Figure A.1.1(a). For further details, see the theory of Gómez, Koon, Lo, Marsden, Masdemont, and Ross [2004].

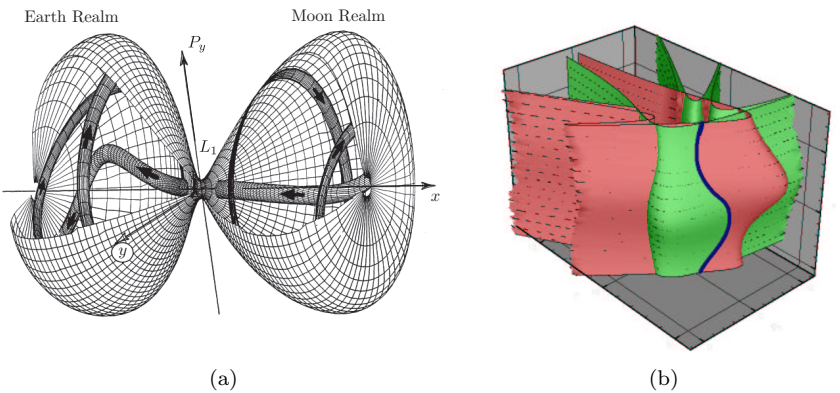


FIGURE A.1.1. **Tube dynamics and lobes dynamics describe the global transport in phase space in a multi-scale way.** (a) Tubes connecting realms, the largest pieces of phase space, as seen in the phase space of a Hamiltonian system with three degrees of freedom. For visualization, three of the phase space dimensions have been suppressed (adapted from Topper [1997]). (b) A 3-dimensional section of the 4-dimensional lobe structure from work by Lekien [2003].

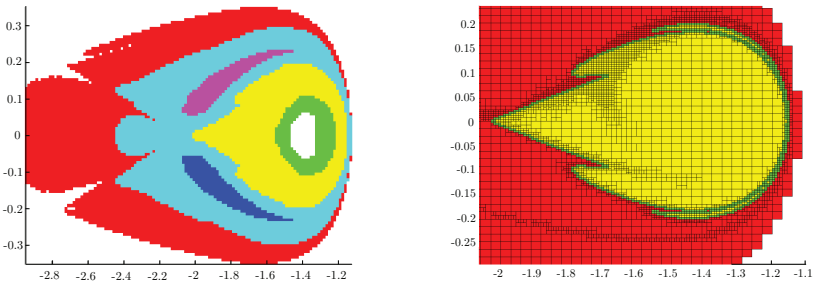
To study the transport between subsets within a particular realm, *lobe dynamics* is also needed. It provides a geometric framework for discussing, describing, and quantifying organized structures relevant to transport within a realm. Tube dynamics has been applied in conjunction with lobe dynamics to give a full picture of the global phase space transport for systems with 2 dof (Koon, Marsden, Ross, Lo, and Scheeres [2004]). This success encourages researchers to extend the combination of the techniques to $n \geq 2$ dof, which preliminary studies have suggested is possible (see Figure A.1.1(b)

and Dellnitz, Junge, Koon, Lekien, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005]; Lekien [2003]).

Merging Tube Dynamics and Lobe Dynamics Into a Single Geometric Theory of Transport. The recent work of Dellnitz, Junge, Koon, Lekien, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005] is a first step toward an exact global theory of transport. Such a picture requires the merging of tube dynamics with lobe dynamics. The main theorem of Koon, Lo, Marsden, and Ross [2000] strongly suggests that tube dynamics and lobe dynamics are linked, and numerical experiments verify this (Koon, Lo, Marsden, and Ross [2001a]; Ross, Koon, Lo, and Marsden [2003, 2004]). Such a unification is strongly needed. Chemists have used variants of both tube dynamics (Topper [1997]; De Leon, Mehta, and Topper [1991a,b]; Ozorio de Almeida, De Leon, Mehta, and Marston [1990]) and lobe dynamics (Davis and Gray [1986]; Davis [1985]) to study molecular dynamics, but these two pictures, which have remained distinct, should come together.

Techniques of Almost Invariant Sets. *Almost invariant set* methods, which use tree structured box elimination and graph partitioning algorithms, have proven to be powerful tools. Developed by the German groups in Paderborn and Berlin (Dellnitz, Froyland, and Junge [2001]; Dellnitz and Hohmann [1997]; Dellnitz, Hohmann, Junge, and Rumpf [1997]; Dellnitz and Junge [1999, 2002]; Dellnitz and Preis [2002]) they have been used to determine, for instance, different molecular conformations and the computation of transition rates between them. The combining of almost invariant set techniques with invariant manifold and lobe dynamics techniques for systems with 2 dof has had a good start. For example, dynamical systems techniques have been used to identify regions in which box refinements are needed and this speeds up the computation considerably. Furthermore, the transport rates using both the box subdivision algorithm and fast and efficient lobe dynamics techniques have been shown to be in excellent agreement (see Figure A.1.2 and Dellnitz, Junge, Koon, Lekien, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005]). This agreement supports the contention that almost invariant set methods will be a powerful tool for studying phase space transport in higher dimensions.

Variational Integration for Accurate Long Term Simulation. For future problems, it will be important to incorporate variational integrators with set oriented methods for the computation of the geometric structures mediating phase space transport in n dof Hamiltonian systems and the associated quantities. *Variational integrators* improve the accuracy of simulations by making use of algorithms based directly on the discretization of variational principles. They provide a systematic and powerful extension of symplectic integrators that have a proven track record in celestial mechanics for long-term integrations (Duncan, Quinn, and Tremaine [1989];



(a) (b)

FIGURE A.1.2. **Almost invariant sets and dynamical systems techniques make computation of phase space transport quantities efficient.** (a) A decomposition of a part of a 2-dimensional Poincaré map into seven almost invariant sets. (b) Suppose we are interested in transport between just two regions. Dynamical systems techniques have been used to identify regions in which box refinements are needed, i.e., where the lobes are located, and this speeds up the computation considerably.

Wisdom and Holman [1991]; Saha and Tremaine [1992]).

Many recent references develop and document the success of this methodology (Marsden, Patrick, and Shkoller [1998]; Kane, Marsden, Ortiz, and West [2000]; Marsden and West [2001]). The exact preservation of conserved quantities, such as energy, is a natural consequence of the discrete variational principle. In addition, many statistical quantities, such as temperature and the structure of chaotic invariant sets, are accurately captured by variational integrators (Lew, Marsden, Ortiz, and West [2003]; Rowley and Marsden [2002]; Lew, Marsden, Ortiz, and West [2004]).

GAIO and MANGEN. Before concluding this section, we give a very brief discussion of the software packages GAIO (Global Analysis of Invariant Objects) and MANGEN (Manifold Generation), which have been used for transport calculations.

GAIO (see `http://www-math.uni-paderborn.de/~agdellnitz/gaio/` and Dellnitz and Junge [2002]) is a set oriented software based on subdivision methods with a searchable tree structured architecture, designed for computing invariant or almost invariant sets in dynamical systems. It also computes transport rates, optimization and related tasks. MANGEN (see Lekien [2003]; Lekien and Coulliette [2003]; Lekien, Coulliette, and Marsden [2003]) is complementary software that computes invariant manifolds, lobe dynamics and related tasks for dynamical systems. It was designed for computing fluid transport properties, such as pollution mixing rates and barriers to transport in time dependent fluid flows. It is quite remarkable that MANGEN and GAIO work together in complementary ways and that MANGEN is also useful in an astrodynamical and dynamical astronomy

context (see Dellnitz, Junge, Koon, Lekien, Lo, Marsden, Padberg, Preis, Ross, and Thiere [2005]).

Software packages like GAIO and MANGEN will be indispensable tools for computational science in the years to come. An important goal of future work is to continue the merging and extension of this software as well as incorporating variational integration techniques.

Bibliography

Abraham, R. and J. E. Marsden [2008], *Foundations of Mechanics*. Originally published in 1967; second edition, 1978; the updated 1985 version reprinted by AMS-Chelsea in 2008, second edition.

Appleyard, D. F. [1970], *Invariant Sets near the Collinear Lagrangian Points of the Nonlinear Restricted Three-Body Problem*, PhD thesis, University of Wisconsin, Madison.

Ascher, U. M. and L. R. Petzold [1998], *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM Publications, Philadelphia.

Barden, B. T. [1994], *Using Stable Manifolds to Generate Transfers in the Circular Restricted Problem of Three Bodies*. M.S. Thesis, Purdue University.

Bate, R. R., D. D. Mueller, and J. E. White [1971], *Fundamentals of Astrodynamics*. Dover, New York.

Belbruno, E. [1994], The Dynamical Mechanism of Ballistic Lunar Capture Transfers in the Four-Body Problem from the Perspective of Invariant Manifolds and Hill's Regions, Preprint no. 270, Center de Recerca Matemàtica, Institut d'Estudis Catalans.

Belbruno, E. [2004], *Capture Dynamics and Chaotic Motions in Celestial Mechanics: With Applications to the Construction of Low Energy Transfers*. Princeton University Press.

Belbruno, E. A. and J. K. Miller [1993], Sun-perturbed Earth-to-Moon transfers with ballistic capture, *Journal of Guidance, Control and Dynamics*, **16**, 770–775.

Bertsekas, D. [1995], *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts.

Bischof, C., A. Carle, G. Corliss, A. Griewank, and P. Hovland [1992], ADIFOR—generating derivative codes from Fortran programs, *Scientific Programming*, **1**, 11–29.

Bischof, C., A. Carle, P. Hovland, P. Khademi, and A. Mauer [1998]. *ADIFOR 2.0 Users' Guide, June 1998.* MCS Division, Argonne National Laboratory, Technical Memorandum No. 192.

Bollt, E. and J. Meiss [1995], Targeting chaotic orbits to the Moon through recurrence, *Physics Letters A*, **204**, 373–378.

Breakwell, J. V. and J. V. Brown [1979], The halo family of 3-dimensional periodic orbits in the Earth-Moon restricted 3-body problem, *Celestial Mechanics*, **20**, 389–404.

Breakwell, J. V., A. A. Kamel, and M. J. Ratner [1974], Station-keeping for a translunar communications station, *Celestial Mechanics*, **10**, 357–373.

Brenan, K. E., S. L. Campbell, and L. R. Petzold [1995], *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations.* SIAM Publications, Philadelphia, 2nd edition.

Burbanks, A., H. Waalkens, and S. Wiggins [2004], Phase space conduits for reaction in multidimensional systems: HCN isomerization in three dimensions, *J. Chem. Phys.*, **121**, 6207–6225.

Chang, D. E., D. F. Chichka, and J. E. Marsden [2002], Lyapunov-based transfer between elliptic Keplerian orbits, *Discrete and Continuous Dynamical Systems–Series B*, **2**, 57–67.

Chang, D. E. and J. E. Marsden [2003], Geometric derivation of the Delaunay variables and geometric phases, *Celestial Mechanics and Dynamical Astronomy*, **86**, 185–208.

Chirikov, B. V. [1979], A universal instability of many-dimensional oscillator systems, *Phys. Rep.*, **52**, 263–379.

Chobotov, V. A. [1996], *Orbital Mechanics.* American Institute of Aeronautics and Astronautics, Reston, Virginia, 2nd edition.

Clarke, A. C. [1950], *Interplanetary Flight.* Temple Press Books Ltd., London.

Condon, G. L. and D. P. Pearson [2001], The role of humans in libration point missions with specific application to an Earth-Moon libration point gateway station. In *AAS/AIAA Astrodynamics Specialist Conference*, Quebec City, Canada. Paper No. AAS 01-307.

Conley, C. C. [1968], Low energy transit orbits in the restricted three-body problem, *SIAM J. Appl. Math.*, **16**, 732–746.

Conley, C. C. [1969], On the ultimate behavior of orbits with respect to an unstable critical point. I. Oscillating, asymptotic, and capture orbits, *J. Differential Equations*, **5**, 136–158.

Ćuk, M. and J. A. Burns [2004], Gas-drag-assisted capture of Himalia's family, *Icarus*, **167**, 369–381.

D'Amario, L. A. [1973], *Minimum Impulse Three-Body Trajectories*, PhD thesis, Massachusetts Institute of Technology.

Davis, M. J. [1985], Bottlenecks to intramolecular energy transfer and the calculation of relaxation rates, *J. Chem. Phys.*, **83**, 1016–1031.

Davis, M. J. and S. K. Gray [1986], Unimolecular reactions and phase space bottlenecks, *J. Chem. Phys.*, **84**, 5389–5411.

De Leon, N. and S. Ling [1994], Simplification of the transition state concept in reactive island theory: Application to the HCN=CNH isomerization, *J. Chem. Phys.*, **101**, 4790–4802.

De Leon, N., M. A. Mehta, and R. Q. Topper [1991], Cylindrical manifolds in phase space as mediators of chemical reaction dynamics and kinetics. I. Theory, *J. Chem. Phys.*, **94**, 8310–8328.

De Leon, N., M. A. Mehta, and R. Q. Topper [1991], Cylindrical manifolds in phase space as mediators of chemical reaction dynamics and kinetics. II. Numerical considerations and applications to models with two degrees of freedom, *J. Chem. Phys.*, **94**, 8329–8341.

De Oliveira, H. P. and I. Damião Soares [1998], Chaos and a resonance mechanism for structure formation in inflationary models, *Modern Physics Letters A*, **13**, 1881–1886.

De Oliveira, H. P., S. L. Sauto, I. Damião Soares, and E. V. Tonini [2001], Chaos and universality in the dynamics of inflationary cosmologies, *Phys. Rev. B*, **60**, 121301–1.

Delaunay, C. [1860], Théorie du Mouvement de la Lune, Mem. **28** (1860); **29** (1867), *acad. sci. france*, Paris, 1860.

Dellnitz, M., G. Froyland, and O. Junge [2001], The algorithms behind GAIO – Set oriented numerical methods for dynamical systems. In Fiedler, B., editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 145–174. Springer, Berlin-Heidelberg-New York.

Dellnitz, M. and A. Hohmann [1997], A subdivision algorithm for the computation of unstable manifolds and global attractors, *Numerische Mathematik* **36**, 491–515.

Dellnitz, M., A. Hohmann, O. Junge, and M. Rumpf [1997], Exploring invariant sets and invariant measures, *Chaos*, **7**, 221.

Dellnitz, M. and O. Junge [1999], On the approximation of complicated dynamical behavior, *SIAM Journal on Numeric Analysis* **36**, 491–515.

Dellnitz, M. and O. Junge [2002], Set oriented numerical methods for dynamical systems. In *Handbook of dynamical systems*, volume 2, pages 221–264. North-Holland, Amsterdam.

Dellnitz, M., O. Junge, W. S. Koon, F. Lekien, M. W. Lo, J. E. Marsden, K. Padberg, R. Preis, S. D. Ross, and B. Thiere [2005], Transport in dynamical astronomy and multibody problems, *International Journal of Bifurcation and Chaos* **15**, 699–727.

Dellnitz, M., O. Junge, M. W. Lo, J. E. Marsden, K. Padberg, R. Preis, S. D. Ross, and B. Thiere [2005], Transport of Mars-crossers from the quasi-Hilda region, *Physical Review Letters* **94**, 231102.

Dellnitz, M. and R. Preis [2002], Congestion and Almost Invariant Sets in Dynamical Systems, In Winkler, F., editor, *Proceedings of SNSC'01*. Springer, Berlin-Heidelberg-New York. to appear.

Deprit, A. [1969], Canonical transformations depending on a small parameter, *Celestial Mechanics*, **1**, 12–30.

Desai, S. D., S. Bhaskaran, W. E. Bollman, C. A. Halsell, J. E. Riedel, and S. P. Synnott [1997], The DS-1 autonomous navigation system: autonomous control of low thrust propulsion system, 1997, Paper No. AAIA-97-3819.

Devaney, R. L. [1981], Singularities in classical mechanical systems. In Katok, A., editor, *Ergodic Theory and Dynamical Systems*, pages 211–333. Birkhäuer, Basel.

Doedel, E. J., R. C. Paffenroth, H. B. Keller, D. J. Dichmann, J. Galan, and A. Vanderbauwhede [2003], Continuation of periodic solutions in conservative systems with application to the 3-body problem, *International Journal of Bifurcation and Chaos*, **13**, 1353–1381.

Duncan, M., T. Quinn, and S. Tremaine [1989], The long-term evolution of orbits in the solar system - A mapping approach, *Icarus*, **82**, 402–418.

Dunn, G. L. [1962], A high-speed data link for farside lunar communications, General Electric Co. Report 62 SPC-5, March 1962.

Euler, L. [1767], De motu rectilineo trium corporum se mutuo attrahentium, *Novi commentarii academiae scientarum Petropolitanae*, **11**, 144–151. In Oeuvres, Seria Secunda tome XXV Commentationes Astronomicae (page 286).

Farquhar, R. W. [1966], Station-keeping in the vicinity of collinear libration points with an application to a Lunar communications problem, in *Space Flight Mechanics, Science and Technology Series*, volume 11, pages 519–535. American Astronautical Society, New York.

Farquhar, R. W. [1969], Future missions for Libration-point satellites, *Astronautics and Aeronautics*, **7**, 52–56.

Farquhar, R. W. [1972], A halo-orbit Lunar station, *Astronautics and Aeronautics*, **10**, 59–63.

Farquhar, R. W. [2001], The flight of ISEE-3/ICE: origins, mission history, and a legacy, *The Journal of the Astronautical Sciences*, **49**, 23–73.

Farquhar, R. W. and D. W. Dunham [1981], A new trajectory concept for exploring the Earth's geomagnetic tail, *Journal of Guidance and Control* **4**, 192–196.

Farquhar, R. W. and A. A. Kamel [1973], Quasi-Periodic orbits about the translunar libration point, *Celestial Mechanics*, **7**, 458–473.

Farquhar, R. W., D. P. Muhonen, C. Newman, and H. Heuberger [1979], The first libration point satellite, mission overview and flight history. In *AAS/AIAA Astrodynamics Specialist Conference*, Provincetown, Massachussetts.

Farquhar, R. W., D. P. Muhonen, C. Newman, and H. Heuberger [1980], Trajectories and orbital maneuvers for the first libration-point satellite, *Journal of Guidance and Control*, **3**, 549–554.

Farquhar, R. W., D. P. Muhonen, and D. L. Richardson [1977], Mission design for a halo orbiter of the Earth, *Journal of Spacecraft and Rockets* **14**, 170–177.

Farquhar, R. [1968], *The control and use of libration-point satellites*, PhD thesis, Stanford University.

Fordy, A. P. [1991], The Hénon-Heiles system revisited, *Physica D* **52**, 201–210.

Gabern, F., W. S. Koon, and J. E. Marsden [2005], Spacecraft dynamics near a binary asteroid, *Discrete Contin. Dyn. Syst.*, 297–306.

Gabern, F., W. S. Koon, J. E. Marsden, and S. D. Ross [2005], Theory and computation of non-RRKM lifetime distributions and rates in chemical systems with three or more degrees of freedom, *Physica D*, **211**, 391–406.

Gabern, F., W. S. Koon, J. E. Marsden, and D. J. Scheeres [2006], Binary asteroid observation orbits from a global dynamical perspective, *SIAM J. Appl. Dyn. Syst.* **5**, 252–279.

Gill, P. E., L. O. Jay, M. W. Leonard, L. R. Petzold, and V. Sharma [2000], An SQP method for the optimal control of large-scale dynamical systems, *J. Comp. Appl. Math.*, **20**, 197–213.

Gill, P. E., W. Murray, and M. A. Saunders [1997]. SNOPT: An SQP algorithm for large-scale constrained optimization, 1997. Numerical Analysis Report 97-2, Department of Mathematics, University of California, San Diego.

Gill, P. E., W. Murray, and M. A. Saunders [1998]. User's guide for SNOPT 5.3: A Fortran package for large-scale nonlinear programming, 1998. Department of Mathematics, University of California, San Diego.

Gill, P. E., W. Murray, and M. H. Wright [1981], *Practical Optimization.* Academic Press, London and New York.

Giorgilli, A., A. Delshams, E. Fontich, L. Galgani, and C. Simó [1989], Effective stability for a Hamiltonian system near an elliptic equilibrium point with an application to the restricted three body problem, *J. Diff. Eq.*, **77**, 167–198.

Gladman, B., J. Burns, M. Duncan, P. Lee, and H. Levison [1996], The exchange of impact ejecta between terrestrial planets, *Science*, **271**, 1387–1392.

Gómez, G., K. Howell, J. Masdemont, and C. Simó [1998], Station keeping strategies for translunar libration point orbits, in *AAS/AIAA Astrodynamics Specialist Conference*, Monterey, California. Paper No. AAS 98-168.

Gómez, G., A. Jorba, J. Masdemont, and C. Simó [1991a], Study refinement of semi-analytical halo orbit theory, Final Report, ESOC Contract No.:8625/89/D/MD(SC), 1991a.

Gómez, G., A. Jorba, J. J. Masdemont, and C. Simó [1991b], A dynamical systems approach for the analysis of the SOHO mission. In *Third International Symposium on Spacecraft Flight Dynamics*, pages 449–454. European Space Agency, Darmstadt, Germany.

Gómez, G., A. Jorba, J. Masdemont, and C. Simó [1993], Study of the transfer from the Earth to a halo orbit around the equilibrium point L_1, *Celestial Mechanics*, **56**, 541–562.

Gómez, G., A. Jorba, J. Masdemont, and C. Simó [1998], Study of the transfer between halo orbits, *Acta Astronautica*, **43**, 493–520.

Gómez, G., J. Llibre, R. Martínez, and C. Simó [2001], *Dynamics and Mission Design near Libration Points – Volume 1: Fundamentals: The Case of Collinear Libration Points*. World Scientific, Singapore.

Gómez, G., J. Llibre, R. Martínez, and C. Simó [2001], *Dynamics and Mission Design near Libration Points – Volume 2: Fundamentals: The Case of Triangular Libration Points*. World Scientific, Singapore.

Gómez, G., A. Jorba, J. Masdemont, and C. Simó [2001], *Dynamics and Mission Design near Libration Points – Volume 3: Advanced Methods for Collinear Points*. World Scientific, Singapore.

Gómez, G., A. Jorba, J. Masdemont, and C. Simó [2001], *Dynamics and Mission Design near Libration Points – Volume 4: Advanced Methods for Triangular Points*. World Scientific, Singapore.

Gómez, G., W. S. Koon, M. W. Lo, J. E. Marsden, J. Masdemont, and S. D. Ross [2001], Invariant manifolds, the spatial three-body problem and space mission design (Paper No. AAS 01-301). In Spencer, D. B., C. C. Seybold, A. K. Misra, and R. J. Lisowski, editors, *Advances in the Astronautical Sciences*, **109**, 3–22. Univelt, San Diego.

Gómez, G., W. S. Koon, M. W. Lo, J. E. Marsden, J. Masdemont, and S. D. Ross [2004], Connecting orbits and invariant manifolds in the spatial three-body problem, *Nonlinearity*, **17**, 1571–1606.

Gómez, G. and J. Masdemont [2000], Some zero cost transfers between libration point orbits, in *AAS/AIAA Space Flight Mechanics Meeting*, Clearwater, Florida. Paper No. AAS 00-177.

Gómez, G., J. Masdemont, and C. Simó [1997], Lissajous orbits around halo orbits, *Advances in the Astronautical Sciences*, **95**, 117–134.

Gómez, G., J. Masdemont, and C. Simó [1998], Quasihalo orbits associated with libration points, *The Journal of the Astronautical Sciences* **46**, 135–176.

Haller, G. [2001], Distinguished material surfaces and coherent structures in 3D fluid flows, *Physica D*, **149**, 248–277.

Haller, G. and I. Mezic [1998], Reduction of three-dimensional, volume-preserving flows with symmetry, *Nonlinearity*, **11**, 319–339.

Haller, G. and G. Yuan [2000], Lagrangian coherent structures and mixing in two-dimensional turbulence, *Physica D*, **147**, 352–370.

Hartman, P. [1964], *Ordinary Differential Equations*. Wiley, New York.

Heaton, A. F., N. J. Strange, J. M. Longuski, and E. P. Bonfiglio [2002], Automated design of the Europa orbiter tour, *Journal of Spacecraft and Rockets*, **39**, 17–22.

Hénon, M. and C. Heiles [1964], The applicability of the third integral of motion: some numerical experiments, *Astron. J.*, **69**, 73–79.

Henrard, J. [1982], Capture into resonance: an extension of the use of adiabatic invariants, *Celest. Mech.*, **27**, 3–22.

Holmes, P. [1990], Poincaré, celestial mechanics, dynamical-systems theory and "chaos", *Physics Reports*, **193**, 137–163.

Howell, K., B. Barden, and M. Lo [1997], Application of dynamical systems theory to trajectory design for a libration point mission, *The Journal of the Astronautical Sciences*, **45**, 161–178.

Howell, K. C. [1984], Three dimensional, periodic, halo orbits, *Celestial Mechanics*, **32**, 53–71.

Howell, K. C., B. T. Barden, R. S. Wilson, and M. W. Lo [1997], Trajectory design using a dynamical systems approach with application to Genesis, in *AAS/AIAA Astrodynamics Specialist Conference*, Sun Valley, Idaho. Paper No. AAS 97-709.

Howell, K. C. and J. V. Breakwell [1984], Almost rectilinear halo orbits, *Celestial Mechanics*, **32**, 29–52.

Howell, K. C. and S. C. Gordon [1994], Orbit determination error analysis and a station-keeping strategy for Sun-Earth L_1 libration point orbits, *The Journal of the Astronautical Sciences*, **42**, 207–228.

Howell, K. C., D. L. Mains, and B. T. Barden [1994], Transfer trajectories from Earth parking orbits to SunEarth halo orbits, in *AAS/AIAA Space Flight Mechanics Meeting*, Cocoa Beach, Florida. Paper No. AAS 94-160.

Howell, K. C. and H. J. Pernicka [1988], Numerical determination of Lissajous trajectories in the restricted three-body problem, *Celestial Mechanics*, **41**, 107–124.

Jaffé, C., D. Farrelly, and T. Uzer [1999], Transition state in atomic physics, *Phys. Rev. A*, **60**, 3833–3850.

Jaffé, C., S. D. Ross, M. W. Lo, J. E. Marsden, D. Farrelly, and T. Uzer [2002], Theory of asteroid escape rates, *Physical Review Letters*, **89**, 011101.

Jorba, A. and J. Masdemont [1999], Dynamics in the Center Manifold of the Collinear Points of the Restricted Three Body Problem, *Physica D* **132**, 189–213.

Kane, C., J. E. Marsden, M. Ortiz, and M. West [2000], Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems, *Int. J. Num. Meth. Eng.*, **49**, 1295–1325.

Karl, D. M., D. F. Bird, K. Bjorkman, T. Houlihan, R. Shackelford, and L. Tupas [1999], Microorganisms in the accreted ice of Lake Vostok, Antarctica, *Science*, **286**, 2144–2147.

Keeter, T. M. [1994], *Station-Keeping Strategies for Libration Point Orbits: Target Point and Floquet Mode Approaches.* M.S. Thesis, Purdue University.

Kirchgraber, U. and D. Stoffer [2004], Possible chaotic motion of comets in the Sun-Jupiter system—a computer-assisted approach based on shadowing, *Nonlinearity*, **17**, 281–300.

Kluever, C. A. [1998], Simple guidance scheme for low-thrust orbit transfer, *Journal of Guidance, Control and Dynamics*, **21**, 1015–1017.

Koon, W. S. and J. E. Marsden [1997], Optimal control for holonomic and nonholonomic mechanical systems with symmetry and Lagrangian reduction, *SIAM J. Control and Optim.*, **35**, 901–929.

Koon, W. S., M. W. Lo, J. E. Marsden, and S. D. Ross [1999], Constructing a Low Energy Transfer between Jovian Moons, in *Celestial Mechanics : an international conference on celestial mechanics*, Evanston, Illinois.

Koon, W. S., M. W. Lo, J. E. Marsden, and S. D. Ross [2000], Heteroclinic connections between periodic orbits and resonance transitions in celestial mechanics, *Chaos*, **10**, 427–469.

Koon, W. S., M. W. Lo, J. E. Marsden, and S. D. Ross [2001a], Resonance and capture of Jupiter comets, *Celestial Mechanics and Dynamical Astronomy*, **81**, 27–38.

Koon, W. S., M. W. Lo, J. E. Marsden, and S. D. Ross [2001b], Low energy transfer to the Moon, *Celestial Mechanics and Dynamical Astronomy*, **81**, 63–73.

Koon, W. S., M. W. Lo, J. E. Marsden, and S. D. Ross [2002], Constructing a low energy transfer between Jovian moons, *Contemporary Mathematics* **292**, 129–145.

Koon, W. S., J. E. Marsden, S. D. Ross, M. W. Lo, and D. J. Scheeres [2004], Geometric mechanics and the dynamics of asteroid pairs, *Annals of the New York Academy of Sciences*, **1017**, 11–38.

Lagrange, J.-L. [1772]. Essai sur le problème des trois corps, 1772. Prix de l'Académie Royale des Sciences de Paris, tome IX, in vol. 6 of *Oeuvres de Lagrange* (Gauthier-Villars, Paris, 1873), 272–282.

Lawden, D. F. [1963], *Optimal Trajectories for Space Navigation.* Butterworth, London.

Lekien, F. [2003], *Time-Dependent Dynamical Systems and Geophysical Flows*, PhD thesis, California Institute of Technology.

Lekien, F. and C. Coulliette [2003], MANGEN: Computation of hyperbolic trajectories, invariant manifolds and lobes of dynamical systems defined as 2D+1 data sets, *in preparation.*

Lekien, F., C. Coulliette, and J. E. Marsden [2003], Lagrangian structures in very high frequency radar data and optimal pollution timing, *7th Experimental Chaos Conference, AIP*, 162–168.

Lekien, F., S. Shadden, and J. Marsden [2007], Lagrangian coherent structures in n-dimensional systems, *J. Math. Physics* **48**, 1–19.

Levison, H. F. and M. J. Duncan [1993], The gravitational sculpting of the Kuiper belt, *Astrophys. J.*, **406**, L35–L38.

Lew, A., J. E. Marsden, M. Ortiz, and M. West [2003], Asynchronous variational integrators, *Archive for Rat. Mech. An.*, **167**, 85–146.

Lew, A., J. E. Marsden, M. Ortiz, and M. West [2004], Variational time integrators, *Intern. J. Num. Meth. in Engin.*, **60**, 153–212.

Li, S. and L. R. Petzold [2000], Software and algorithms for sensitivity analysis of large-scale differential algebraic systems, *J. Comp. Appl. Math.*, **125**, 131–145.

Liu, Y., K. L. Teo, L. S. Jennings, and S. Wang [1998], On a Class of Optimal Control Problems with State Jumps, *J. of Optim. Theory and Appl.* **98**, 65–82.

Llibre, J., R. Martinez, and C. Simó [1985], Transversality of the invariant manifolds associated to the Lyapunov family of periodic orbits near L2 in the restricted three-body problem, *J. Diff. Eqns.*, **58**, 104–156.

Lo, M., B. G. Williams, W. E. Bollman, D. Han, Y. Hahn, J. L. Bell, E. A. Hirst, R. A. Corwin, P. E. Hong, K. C. Howell, B. Barden, and R. Wilson [2001], Genesis mission design, *The Journal of the Astronautical Sciences*, **49**, 169–184.

Lo, M. W. and S. D. Ross [1998], Low energy interplanetary transfers using invariant manifolds of L1, L2 and halo orbits, in *AAS/AIAA Space Flight Mechanics Meeting*, Monterey, California.

Lo, M. W. and S. D. Ross [2001], The Lunar L1 Gateway: Portal to the stars and beyond, in *AIAA Space 2001 Conference*, Albequerque, New Mexico.

Ludwinski, J. M., M. D. Guman, J. R. Johannesen, R. T. Mitchell, and R. L. Staehle [1998], The Europa Orbiter Mission Design, in *49th International Astronautical Congress*, Melbourne, Australia, Paper No. 98-4.2.02.

Luus, R. [2000], *Iterative Dynamic Programming.* CRC Press, London.

MacKay, R. S., J. D. Meiss, and I. C. Percival [1984], Transport in Hamiltonian systems, *Physica D*, **13**, 55–81.

MacKay, R. S., J. D. Meiss, and I. C. Percival [1987], Resonances in area-preserving maps, *Physica D*, **27**, 1–20.

Mains, D. L. [1993], *Transfer Trajectories from Earth Parking Orbits to L_1 Halo Orbits*, M.S. Thesis, Purdue University.

Malhotra, R. [1996], The phase space structure near Neptune resonances in the Kuiper belt, *Astron. J.*, **111**, 504–516.

Malhotra, R., M. Duncan, and H. Levison [2000], Dynamics of the Kuiper belt. In Mannings, V., A. P. Boss, and S. S. Russell, editors, *Protostars and Planets IV*, pages 1231–1254. Univ. of Arizona Press, Tucson.

Marchal, C. [1990], *The Three-Body Problem.* Elsevier, Amsterdam.

Marsden, J. E., G. W. Patrick, and S. Shkoller [1998], Multisymplectic geometry, variational integrators and nonlinear PDEs, *Comm. Math. Phys.*, 351–395.

Marsden, J. E. and T. S. Ratiu [1999], *Introduction to Mechanics and Symmetry*, volume 17 of *Texts in Applied Mathematics.* Springer-Verlag.

Marsden, J. E. and S. D. Ross [2006], New methods in celestial mechanics and mission design, *Bull. Amer. Math. Soc. (N.S.)* **43**, 43–73.

Marsden, J. E. and J. Scheurle [1995], Pattern evocation and geometric phases in mechanical systems with symmetry, *Dynamics Stability Systems* **10**, 315–338.

Marsden, J. E. and M. West [2001], Discrete mechanics and variational integrators, *Acta Numerica*, **10**, 357–514.

Martens, C. C., M. J. Davis, and G. S. Ezra [1987], Local frequency analysis of chaotic motion in multidimensional systems: energy transport and bottlenecks in planar OCS, *Chem. Phys. Lett.*, **142**, 519–528.

McGehee, R. [1969], *Some homoclinic orbits for the restricted three-body problem*, PhD thesis, University of Wisconsin, Madison.

Meiss, J. D. [1992], Symplectic maps, variational principles, and transport, *Rev. Mod. Phys.*, **64**, 795–848.

Meiss, J. D. and E. Ott [1986], Markov tree model of transport in area-preserving maps, *Physica D*, **20**, 387–402.

Meyer, K. R. and R. Hall [1992], *Hamiltonian Mechanics and the N-Body Problem*, Texts in Applied Mathematics Science. Springer-Verlag, Berlin.

Milam, M. B. [2003], *Real-time optimal trajectory generation for constrained dynamical systems*, PhD thesis, California Institute of Technology.

Moser, J. [1958], On the generalization of a theorem of Liapunov, *Comm. Pure Appl. Math.*, **11**, 257–271.

Moser, J. [1973], *Stable and Random Motions in Dynamical Systems with Special Emphasis on Celestial Mechanics*. Princeton University Press.

Murray, N. and M. Holman [2001], The role of chaotic resonances in the Solar System, *Nature*, **410**, 773–779.

Neishtadt, A. [1996], Scattering by resonances, *Celest. Mech. Dyn. Astr.* **65**, 1–20.

Neishtadt, A. I., V. V. Sidorenko, and D. V. Treschev [1997], Stable periodic motions in the problem on passage through a separatrix, *Chaos* **7**, 2–11.

Ozorio de Almeida, A. M., N. De Leon, M. A. Mehta, and C. C. Marston [1990], Geometry and dynamics of stable and unstable cylinders in Hamiltonian systems, *Physica D*, **46**, 265–285.

Parker, T. S. and L. O. Chua [1989], *Practical Numerical Algorithms for Chaotic Systems*. Springer-Verlag, New York.

Pierrehumbert, R. T. [1991], Chaotic mixing of tracer and vorticity by modulated traveling Rossby waves, *Geophys. Astrophys. Fluid Dyn.* **58**, 285–319.

Pierrehumbert, R. T. [1991], Large-scale horizontal mixing in planetary atmospheres, *Phys. Fluids A*, **3**, 1250–1260.

Poincaré, H. [1890], Sur la problème des trois corps et les équations de la dynamique, *Acta Math.*, **13**, 1–271.

Poje, A. C. and G. Haller [1999], Geometry of cross-stream mixing in a double-gyre ocean model, *Phys. Oceanogr.*, **29**, 1649–1665.

Rehbock, V., K. L. Teo, L. S. Jennings, and H. W. J. Lee [1999], A Survey of the control parametrization and control parametrization enhancing methods for constrained optimal control, in Eberhard, A., R. Hill, D. Ralph, and B. M. Glover, editors, *Progress in Optimization; Contributions from Australia*. Kluwer Academic.

Richardson, D. L. [1980], Analytical construction of periodic orbits about the collinear points, *Celestial Mechanics*, **22**, 241–253.

Richardson, D. L. [1980], Halo orbit formulation for the ISEE-3 mission, *Journal of Guidance and Control*, **3**, 543–548.

Rom-Kedar, V. and S. Wiggins [1990], Transport in two-dimensional maps, *Arch. Rat. Mech. Anal.*, **109**, 239–298.

Ross, S. D. [2003], Statistical theory of interior-exterior transition and collision probabilities for minor bodies in the solar system. In Gómez, G., M. W. Lo, and J. J. Masdemont, editors, *Libration Point Orbits and Applications*, pages 637–652. World Scientific.

Ross, S. D. [2004], *Cylindrical manifolds and tube dynamics in the restricted three-body problem*, PhD thesis, California Institute of Technology.

Ross, S. D., W. S. Koon, M. W. Lo, and J. E. Marsden [2003], Design of a multi-moon orbiter, in *13th AAS/AIAA Space Flight Mechanics Meeting*, Ponce, Puerto Rico, Paper No. AAS 03-143.

Ross, S. D., W. S. Koon, M. W. Lo, and J. E. Marsden [2004], Application of dynamical systems theory to a very low energy transfer, in *14th AAS/AIAA Space Flight Mechanics Meeting*, Maui, Hawaii, Paper No. AAS 04-289.

Ross, S. D. and D. J. Scheeres [2007], Multiple gravity assists, capture, and escape in the restricted three-body problem, *SIAM J. Appl. Dyn. Syst.* **6**, 576–596.

Rowley, C. W. and J. E. Marsden [2002], Variational integrators for point vortices, *Proc. CDC*, **40**, 1521–1527.

Roy, A. E. [1988], *Orbital Motion*. Adam Hilger, Bristol, 3rd edition.

Saha, P. and S. Tremaine [1992], Symplectic integrators for solar system dynamics, *Astron. J.*, **104**, 1633–1640.

Scheeres, D. J., M. D. Guman, and B. Villac [2001], Stability analysis of planetary satellite orbiters: Application to the Europa Orbiter, *Journal of Guidance, Control and Dynamics*, **24**, 778–787.

Schoenmaekers, J., D. Horas, and J. A. Pulido [2001], SMART-1: With solar electric propulsion to the Moon, in *16th International Symposium on Space Flight Dynamics*, Pasadena, CA, Dec. 2001.

Schroer, C. G. and E. Ott [1997], Targeting in Hamiltonian systems that have mixed regular/chaotic phase spaces, *Chaos*, **7**, 512–519.

Serban, R., W. Koon, M. Lo, J. Marsden, L. Petzold, S. Ross, and R. Wilson [2002], Halo orbit mission correction maneuvers using optimal control, *Automatica*, **38**, 571–583.

Serban, R. and L. R. Petzold [2001], COOPT - A software package for optimal control of large-scale differential-algebraic equation systems, *Mathematics and Computers in Simulation*, **56**, 187–203.

Shadden, S. C., F. Lekien, and J. E. Marsden [2005], Definition and properties of Lagrangian coherent structures: Mixing and transport in two-dimensional aperiodic flows, *Physica D* **212**, 271–304.

Simó, C. [1989], Estabilitat de sistemes Hamiltonians, *Mem. Real Acad. Cienc. Artes Barcelona*, **48**(7), 303–348.

Simó, C. [1999], Dynamical systems methods for space missions on a vicinity of collinear libration points, in Simó, C., editor, *Hamiltonian Systems with Three or More Degrees of Freedom (S'Agaró, 1995)*, volume 533 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 223–241, Dordrecht. Kluwer Acad. Publ.

Simó, C., G. Gómez, A. Jorba, and J. Masdemont [1995], The bicircular model near the triangular libration points, in Roy, A. E. and B. A. Steves, editors, *From Newton to Chaos*, pages 343–370. Plenum Press, New York.

Simó, C., G. Gómez, J. Llibre, and R. Martinez [1986], Station keeping of a quasiperiodic halo orbit using invariant manifolds, in *Second International Symposium on Spacecraft Flight Dynamics*, pages 65–70. European Space Agency, Darmstadt, Germany.

Simó, C., G. Gómez, J. Llibre, R. Martinez, and R. Rodríguez [1987], On the optimal station keeping control of halo orbits, *Acta Astronautica* **15**, 391–397.

Stalos, S., D. Folta, B. Short, J. Jen, and A. Seacord [1993], Optimum transfer to a large-amplitude halo orbit for the Solar and Heliospheric Observatory (SOHO) spacecraft, *Advances in the Astronautical Sciences*, **84**, 639–650.

Sweetser, T., R. Maddock, J. Johannesen, J. Bell, P. Penzo, A. Wolf, S. Williams, S. Matousek, and S. Weinstein [1997], Trajectory design for a Europa orbiter mission: A plethora of astrodynamic challenges, in

AAS/AIAA Space Flight Mechanics Meeting, Huntsville, Alabama, Paper No. AAS 97-174.

Szebehely, V. [1967], *Theory of Orbits: The Restricted Problem of Three Bodies*, Academic, New York.

Topper, R. Q. [1997], Visualizing molecular phase space: nonstatistical effects in reaction dynamics, in Lipkowitz, K. and D. Boyd, editors, *Reviews in Computational Chemistry*, volume 10, chapter 3, pages 101–176. VCH Publishers, New York.

Topputo, F., V. Massimiliano, and A. E. Finzi [2004], An approach to the design of low energy interplanetary transfers exploiting invariant manifolds of the restricted three-body problem, in *14th AAS/AIAA Space Flight Mechanics Meeting*, Maui, Hawaii. Paper No. AAS 04-245.

Torbett, M. V. and R. Smoluchowski [1990], Chaotic motion in a primordial comet disk beyond Neptune and comet influx to the Solar System, *Nature* **345**, 49–51.

Vainchtein, D. and I. Mezić [2004], Capture into resonance: a method for efficient control, *Physical Review Letters*, **93**, 084301.

Valsecchi, G. B. and A. Manara [1997], Dynamics of comets in the outer planetary region. II. Enhanced planetary masses and orbital evolutionary paths, *Astron. Astophys.*, **323**, 986–998.

Vela-Arevalo, L. [2002], *Time-frequency analysis based on wavelets for Hamiltonian systems*, PhD thesis, California Institute of Technology.

Vela-Arevalo, L. and J. E. Marsden [2004], Time-frequency analysis of the restricted three body problem: Transport and resonance transitions, *Classical and Quantum Gravity* **21**, S351–S375.

Villac, B. F. and D. J. Scheeres [2001], Escaping trajectories in the Hill three-body problem and applications, *Journal of Guidance, Control and Dynamics*, **26**, 224–232.

Whiffen, G. J. [2003], An investigation of a Jupiter Galilean moon orbiter trajectory, in *AAS/AIAA Astrodynamics Specialist Conference*, Big Sky, Montana, Paper No. AAS 03-544.

Whittaker, E. T. [1927], *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies: With an Introduction to the Problem of Three Bodies*, Cambridge University Press, Cambridge.

Wiesenfeld, L., A. Faure, and T. Johann [2003], Rotational transition states: relative equilibrium points in inelastic molecular collisions, *J. Phys. B: At. Mol. Opt. Phys.*, **36**, 1319–1335.

Wiggins, S. [1993], *Global Dynamics, Phase Space Transport, Orbits Homoclinic to Resonances, and Applications.* Fields Institute Monographs. American Mathematical Society.

Wiggins, S. [1994], *Normally Hyperbolic Invariant Manifolds in Dynamical Systems*, Springer-Verlag, New York.

Wiggins, S. [2003], *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, volume 2 of *Texts in Applied Mathematics Science.* Springer-Verlag, Berlin, 2nd edition.

Wiggins, S., L. Wiesenfeld, C. Jaffé, and T. Uzer [2001], Impenetrable barriers in phase space, *Physical Review Letters*, **86**, 5478–5481.

Wilczak, S. and P. Zgliczyński [2003], Heteroclinic connections between periodic orbits in planar restricted circular three body problem—a computer assisted proof, *Commun. Math. Phys.*, **234**, 37–75.

Wilson, R. S. and K. C. Howell [1998], Trajectory design in the Sun-Earth-Moon system using Lunar gravity assists, *Journal of Spacecraft and Rockets*, **35**, 191–198.

Wilson, R. S., K. C. Howell, and M. W. Lo [1999], Optimization of insertion cost for transfer trajectories to libration point orbits. In *AAS/AIAA Astrodynamics Specialist Conference*, Girdwood, Alaska, Paper No. AAS 99-401.

Wisdom, J. [1980], The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body, *Astron. J.*, **85**, 1122–1133.

Wisdom, J. and M. Holman [1991], Symplectic maps for the n-body problem, *Astron. J.*, **102**, 1528–1538.

Yamato, H. and D. B. Spencer [2003], Numerical investigation of perturbation effects on orbital classifications in the restricted three-body problem. In *AAS/AIAA Space Flight Mechanics Meeting*, Ponce, Puerto Rico. Paper No. AAS 03-235.

Zondervan, K. P., L. J. Wood, and T. K. Caughy [1984], Optimal low-thrust, three-burn transfers with large plane changes, *The Journal of the Astronautical Sciences*, **32**, 407–427.

Index