

CHAPTER 2

METHODOLOGY

2.1 METHODOLOGY INTRODUCTION

This chapter introduces all of the models, coordinate frames, and methodology used in the analysis and construction of lunar transfers. The chapter begins by simply defining the physical constants used in these analyses, including the masses and radii of the Sun, the Moon, and the planets. It then defines the time systems used, coordinate frames, and models, including the circular restricted three-body problem and the Jet Propulsion Laboratory (JPL) developmental ephemerides used to model the motion of the planets and the Moon. A large portion of this chapter is then devoted to describing the dynamical systems methods employed in this work for the analysis and design of low-energy transfers in the Solar System. These methods may be used to design low-energy transfers from one orbit to another and/or one celestial body to another, such as low-energy transfers between the Earth and the Moon. Finally, this chapter discusses the tools used to generate the trajectories in this work.

2.2 PHYSICAL DATA

The trajectories generated in this work have been propagated using point masses for the Sun, the Moon, and the planets. Early analyses include just the Sun, Earth, and Moon, often in circular orbits that approximate the real orbits. Once early analyses are complete, high-fidelity trajectories are generated that include all of the planets, such that their positions are determined at each moment in time using accurate planetary ephemerides. Table 2-1 presents the masses, gravitational parameters, and average radii used to generate each trajectory, where it is assumed that the gravitational constant, G is equal to 6.673×10^{-20} cubic kilometers per second squared per kilogram ($\text{km}^3/\text{s}^2/\text{kg}$).

The values of gravitational constant times mass (GM) shown in cubic kilometers per second squared (km^3/s^2) in Table 2-1 are the best estimates of those values when modeling the entire Solar System as point masses. However, other GM values represent the best estimate for different cases, such as when one is modeling the gravity of the Moon using the spherical harmonic expansion. For instance, the LP150Q gravity field estimates the GM of the Moon to be approximately 4902.801076 km^3/s^2 : slightly different than the value in the table [88].

Table 2-1 The masses, gravitational parameters, and average radii of the Sun, Moon, and planets used in this work [89, 90]. If the planet has natural satellites, the mass and gravitational parameter of the barycenter of the planetary system have been used.

Body	Mass (kg)	GM (km^3/s^2)	Radius (km)
Sun	$1.98879724 \times 10^{30}$	$1.32712440 \times 10^{11}$	696000.
Earth	$5.97333183 \times 10^{24}$	3.98600433×10^5	6378.14
Moon	$7.34722101 \times 10^{22}$	4.90280058×10^3	1737.4
Earth Barycenter	$6.04680404 \times 10^{24}$	4.03503233×10^5	–
Mercury	$3.30167548 \times 10^{23}$	2.20320805×10^4	2439.7
Venus	$4.86825414 \times 10^{24}$	3.24858599×10^5	6051.8
Mars	$6.41814926 \times 10^{23}$	4.28283100×10^4	3396.19
Mars Barycenter	$6.41814990 \times 10^{23}$	4.28283143×10^4	–
Jupiter	$1.89849445 \times 10^{27}$	1.26686534×10^8	71492.
Jupiter Barycenter	$1.89888757 \times 10^{27}$	1.26712768×10^8	–
Saturn	$5.68552375 \times 10^{26}$	3.79395000×10^7	60268.
Saturn Barycenter	$5.68569250 \times 10^{26}$	3.79406261×10^7	–
Uranus	$8.68269993 \times 10^{25}$	5.79396566×10^6	25559.
Uranus Barycenter	$8.68357412 \times 10^{25}$	5.79454901×10^6	–
Neptune	$1.02429180 \times 10^{26}$	6.83509920×10^6	24764.
Neptune Barycenter	$1.02450683 \times 10^{26}$	6.83653406×10^6	–
Pluto	$1.32300764 \times 10^{22}$	8.82843000×10^2	1195.
Pluto Barycenter	$1.47100388 \times 10^{22}$	9.81600888×10^2	–

The radius of the Earth at the Equator is equal to approximately 6378.14 km, according to the International Astronomical Union/International Association of Geodesy (IAU/IAG) 2000 Report [89]. The distance from the Earth's center to either pole is approximately 6356.75 km, shorter than at the Equator since the Earth has a significant oblateness about the Equator [89]. The radius that defines the atmospheric boundary at the Earth for sample return missions is equal to approximately 6503.14 km, approximately 125 km above the Earth's Equator [91].

2.3 TIME SYSTEMS

The passage of time may be represented in countless ways. One may define broad definitions of four types of time systems that are in common use in physics and astronomy. To varying degrees, each of these types of time systems, and the relationships between them, is important to the mission analyst [91, 92].

1. Dynamical time, in which the unit of duration is based on the orbital motion of the Earth, Moon, and planets.
2. Atomic time, in which the unit of duration corresponds to a defined number of wavelengths of radiation of a specified atomic transition of a chosen isotope.
3. Universal time, in which the unit of duration represents the solar day, defined to be as uniform as possible, despite variations in the rotation of the Earth.
4. Sidereal time, in which the unit of duration is the period of the Earth's rotation with respect to a point nearly fixed with respect to the stars.

It is very difficult to be both succinct and technically correct when defining the different types of time systems that exist. See Seidelmann, 1992, for more details [92].

2.3.1 Dynamical Time, ET

To a mission analyst, “ephemeris time” or “ET” refers to the independent variable in the equations of motion governing the motion of bodies in the Solar System. The time scale represents a smooth-flowing time coordinate that is used in the development of the numerically integrated Solar System ephemerides produced at JPL and distributed worldwide [91], as well as barycentric dynamical time (TDB). This time scale has also been referred to as T_{eph} in other studies [93]. Unfortunately, the label “ET” has a history of referring to a variety of slightly different time scales in previous studies.

2.3.2 International Atomic Time, TAI

As of 2012, the fundamental time period of a *second* is defined in the Système International (SI) system to be a specific number of oscillations of an undisturbed

cesium atom. Specifically, the second is defined as the duration of time required for 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom. The Temps Atomique International (TAI), or international atomic time, is defined as a continuous time scale resulting from the statistical analysis of a large number of atomic clocks operating around the world, performed by the Bureau International des Poids et Mesures (BIPM). The difference between Terrestrial Time (TT) and TAI is approximately 32.184 seconds (s); that is, $TT - TAI = 32.184 \text{ s}$. The difference between TAI and ET is: $ET - TAI = 32.184 \text{ s} + \text{relativistic terms}$, where the relativistic terms contribute less than 2 milliseconds (ms) of variation [91].

2.3.3 Universal Time, UT

Universal Time (UT) is a time scale that is based upon the mean solar day. The time scale “UT1” represents the daily rotation of the Earth and is independent of the observing location, that is, it is independent of corrections for polar motion on the longitude of the observing site. The Earth’s rotation rate changes continuously as its shape and mass distribution shifts; hence, this time scale is unpredictable. UT1 is computed using a combination of a variety of different types of observations, including very long baseline interferometry (VLBI) measurements of extragalactic radio sources (quasars), lunar laser ranging, satellite laser ranging, and Global Positioning System (GPS) measurements, to name a few.

2.3.4 Coordinated Universal Time, UTC

The Coordinated Universal Time (UTC) is the time scale that is used as the basis for the worldwide system of civil timekeeping and is available from radio broadcast signals. It is the time system used by flight operations teams and tracking stations. UTC was set equal to TAI in 1958; it was reset in 1972 such that the TAI time scale was 10 s ahead of UTC, corresponding to the approximate accumulation of drift by 1972. From then on it has been adjusted using leap seconds so that it remains within 0.9 s of UT1. As of early 2012, a total of 24 leap seconds had been added, such that the TAI time scale was 34 s ahead of UTC, that is, $TAI - UTC = 34 \text{ leap seconds}$. The “ET” time scale was 66.184 s (excluding periodic relativistic terms) ahead of UTC, as it had been since January 1, 2009 [91].

2.3.5 Lunar Time

A “day” on the Moon is typically associated with a mean solar day, namely, the duration of time between sunrises and sunsets at a particular location on the surface. Put another way, a day on the Moon is equal to the mean interval of time between successive crossings of the Sun on a particular lunar longitude, that is, the lunar prime meridian. As a result, the period of one mean lunar day is equal to the period of a mean synodic lunar month, namely, approximately 29.53059 Earth days. The actual lunar month may vary from this mean value by nearly ± 2 hours due to the

eccentricity of the Earth's orbit and small periodic variations in the Moon's rotation rate.

A lunar month may be defined in a variety of ways. Table 2-2 summarizes five ways that one may define a lunar month and their corresponding durations of time [92].

2.3.6 Local True Solar Time, LTST

The Local True Solar Time (LTST) represents the instantaneous time of day of an observer at a site on the Moon. It is a time system that does not flow constantly, but it is useful to the mission planner when measuring time for a lunar lander. The LTST on the Moon is defined as follows

$$\text{LTST} = (\lambda_p - \lambda_{TS} \text{ deg}) \frac{24 \text{ hr}}{360 \text{ deg}} + 12 \text{ hr}$$

where λ_p is the east longitude of a point on the surface of the Moon and λ_{TS} is the east longitude of the true Sun. Using this relationship, 12 lunar hours corresponds to the time when the Sun is crossing the local meridian of the reference site, for example, local noon, and the lunar day includes 24 lunar hours.

2.3.7 Orbit Local Solar Time, OLST

During the development and operations of nearly all planetary and satellite orbiting missions, understanding how the geometry of the orbit plane changes relative to the Sun over time is extremely important, both from an engineering and a science perspective. A useful way to characterize the orbit geometry, particularly for high-inclination orbiters, is to report the local solar time of the ascending or descending node of the orbit, namely, the Orbit Local Solar Time (OLST). To be clear, this measurement describes the orientation of the orbit relative to the Moon's surface. The convention generally adopted is to report the local time of the orbit node relative to the true Sun. The Moon's gravity field will have an effect on the orbit's OLST

Table 2-2 Five ways to define a lunar month and their corresponding durations of Earth time [92].

Month	Duration	Duration			
	(Earth days)	days	hr	min	s
Synodic (new moon to new moon)	29.53059	29	12	44	03
Anomalistic (perigee to perigee)	27.55455	27	13	18	33
Sidereal (fixed star to fixed star)	27.32166	27	07	43	12
Tropical (equinox to equinox)	27.32158	27	07	43	05
Nodical / Draconic (node to node)	27.21222	27	05	05	36

over time, but to first order the orbit remains essentially fixed in inertial space. The main reason that the orbit's OLST will change over time is due to the motion of the Earth–Moon system about the Sun. Thus, the following relationship describes the gross change in the OLST over time, derived from the mean synodic and sidereal periods of the Moon's orbit.

$$\text{OLST}_{\text{in LTST}} \text{ changes by } -3.94 \frac{\text{minutes}}{\text{Earth day}} = -27.60 \frac{\text{minutes}}{\text{week}} = -1.94 \frac{\text{hours}}{\text{mean lunar day}}$$

Since the change in OLST over time is primarily a function of the rate at which the Earth–Moon system moves about the Sun, the partial will change slightly as a function of time due to the eccentricity of the Earth's orbit. For example, the partial will vary roughly within the following range each year during the 3-year period from 2009–2012

$$\text{OLST}_{\text{in LTST}} \text{ changes by } -4.1 \frac{\text{minutes}}{\text{Earth day}} \text{ to } -3.8 \frac{\text{minutes}}{\text{Earth day}} \text{ within 2009–2012}$$

2.4 COORDINATE FRAMES

This section describes several coordinate frames that are frequently used in lunar mission analysis. Each coordinate frame has its use: some are useful to describe states on the surface of the Earth, Moon, or other body; others are useful to describe the relative geometry between the Sun, Earth, and/or Moon.

Coordinate systems include a reference frame and an origin, and are often rotating or translating relative to other bodies. A coordinate system is inertial only when it is not accelerating. When referencing motion in the Solar System, the only truly “inertial” coordinate system is one that is not rotating and centered at the Solar System barycenter. Strictly speaking, no Earth-centered coordinate system can be inertial, even one that is not rotating, since the Earth is accelerating in its orbit as it revolves about the Sun. Although it is inaccurate, coordinate systems may be referred to in this book as “inertial” when they are merely nonrotating.

2.4.1 EME2000

The Earth Mean Equator and Equinox of J2000 (EME2000) coordinate frame is a nonrotating coordinate frame that is approximately aligned with the Earth's Equator. It is almost identical to the International Celestial Reference Frame (ICRF) [94]. The ICRF is defined by the IAU and is tied to the observations of a selection of quasars and other distant bright radio objects. It is a reference frame that is fixed as well as possible to the observable universe. Each of the quasars moves relative to the others, but very slowly; the motion of each of the quasars is averaged out in order to best approximate inertial space relative to the Earth's position in the universe. According to Feissel and Mignard [95], the pole of the EME2000 frame differs from the ICRF pole by ~ 18 milliarcseconds and the right ascension of the EME2000 x -axis differs from the right ascension of the ICRF x -axis by 78 milliarcseconds.

The coordinate axes are defined as follows:

- The z -axis of the EME2000 is defined as the pole vector of the Earth Mean Equator at the J2000 epoch, namely, at 1 January 2000 12:00:00 ET, or at Julian date 2451545.0 ET.
- The x -axis of the EME2000 is defined as the cross product of the z -axis and the Earth Mean Orbit pole of J2000, that is, the ecliptic pole of J2000. This axis defines the vernal equinox of J2000.
- The y -axis completes the right-handed coordinate frame.

This coordinate frame provides the fundamental reference for the definitions of other coordinate frames.

2.4.2 EMO2000

The Earth Mean Orbit of J2000 (EMO2000) coordinate frame is a nonrotating coordinate frame that is approximately aligned with the ecliptic. The frame shares the same x -axis as the EME2000 frame, but is rotated about that axis such that the EMO2000 z -axis is aligned with the mean ecliptic pole of J2000. This involves a rotation of approximately 23.4393 degrees (deg). The y -axis completes the right-handed coordinate frame.

2.4.3 Principal Axis Frame

The principal axis frame of a body is a body-fixed coordinate frame, that is, rotating frame, aligned with the principal axes of the body.

The coordinate axes are defined as follows:

- The z -axis points in the direction of the maximum moment of inertia; for the Earth and the Moon, this is the North Pole principal axis.
- The x -axis points in the direction of the minimum moment of inertia, that is, the prime meridian principal axis.
- The y -axis completes the right-handed coordinate frame.

It is common practice to define lunar gravity fields in the lunar principal-axis body-fixed frame (LPABF).

2.4.4 IAU Frames

The International Astronomical Union has developed definitions for coordinate frames that are tied to the surface of each planet, many satellites, and some other bodies in the Solar System. There are typically two variations of each coordinate frame: a fixed frame that rotates with the motion of the body about its primary spin axis and an “inertial” frame that shares the same z -axis but which does not rotate.

Detailed definitions of the IAU frames are described by Archinal et al. [96], and an overview of these frames is given here. The z -axis of the IAU body-fixed frame for a given body is aligned with the direction of the spin-axis of that body. The positive direction of the North Pole is defined to be on the north side of the invariable plane of the Solar System (defined by the angular momentum of the Solar System), and the pole's orientation is defined using measured values for the right ascension and declination [96]. Relatively simple low-degree polynomial approximations are used to compute the direction of this pole vector for most of the planets. Longitude is typically defined relative to a fixed surface feature for rigid bodies. In each case these quantities are specified relative to the ICRF, which varies slightly from the EME2000 coordinate frame as described above.

To give some idea for the variations between the Earth's IAU frame and EME2000, Fig. 2-1 illustrates the mapping of Greenwich, England, from the inertial IAU Earth frame to EME2000, where Greenwich has been defined in the IAU Earth frame to be at a latitude of 51.48 deg North and a longitude of 0.0 deg at the J2000 epoch.

2.4.5 Synodic Frames

It is often useful to describe a synodic frame that rotates with the motion of two massive bodies about their barycenter. Two synodic reference frames that are frequently used in this work are the Earth–Moon synodic frame, which rotates with the motion of the Earth and the Moon about their barycenter, and the Sun–Earth synodic frame, which rotates with the motion of the Earth–Moon barycenter about the Sun. The

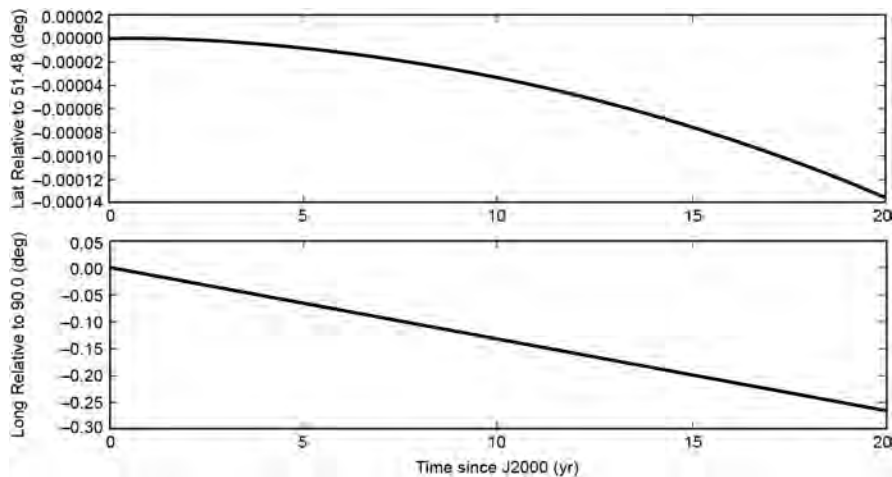


Figure 2-1 The latitude and longitude of Greenwich, England, in EME2000, where Greenwich has been defined in the inertial IAU Earth frame to be at a latitude of 51.48 deg North and a longitude of 0.0 deg at the J2000 epoch.

synodic frame may be constructed to rotate at a constant rate or at a rate that varies with the instantaneous motion of the bodies. In this book, the frames are always constructed such that the x -axis points from the larger body to the smaller body at each instant in time, the z -axis points in the direction of the angular momentum of the system, and the y -axis completes the right-handed coordinate system. This definition defines a frame that rotates at a rate that varies with the motion of the bodies in their orbits. Of course, if the bodies orbit their barycenter in circular orbits, then this frame rotates at a constant rate.

2.5 MODELS

This section describes the different models that have been used in this work to approximate the motion of spacecraft in the Solar System. Each model has a use in the analyses provided here.

The most basic model is the *two-body model*, which is used to approximate the motion of a spacecraft about a simple massive body without any other perturbations. This model is useful because one can use conic sections to approximate the spacecraft's motion, which are predictable and very quick to generate. This model is very well-known [97] and will not be further described here.

The next step up in complexity is a model that includes the gravitational attraction of two large bodies, namely, the model formulated by the circular restricted three-body problem (CRTBP). The CRTBP more closely approximates the motion of a spacecraft in the Earth–Moon and Sun–Earth three-body systems than the two-body model. Working within the CRTBP allows a mission designer to bring a wealth of techniques that have been developed over hundreds of years to a design problem. These techniques provide many qualitative insights that assist in the design of useful low-energy orbit transfers.

The patched three-body model gracefully introduces a fourth body into the design problem. The patched three-body model approximates the motion of a spacecraft using the Sun–Earth three-body model for all times, except when the spacecraft is within close proximity to the Moon, at which point the model approximates the motion of the spacecraft using the Earth–Moon three-body model. These features permit the design of four-body trajectories, such as low-energy lunar transfers, while retaining much of the useful structure found in the CRTBP.

Finally, the fourth and most complex model frequently used in this work is the JPL developmental ephemerides model, which approximates the motion of a spacecraft under the influence of the gravitational attraction of any or all of the planets and the Moon, using accurate ephemerides to model the motion of the planets and the Moon relative to the Sun.

Each of these models is described in detail in this section.

2.5.1 CRTBP

The CRTBP describes a dynamical model that is used to characterize the motion of a massless particle, for example, a spacecraft, in the presence of two massive bodies, such as the Earth and the Moon [86]. The model assumes the two massive bodies orbit their barycenter in circular orbits.

2.5.1.1 Equations of Motion It is convenient to characterize the motion of the third body, that is, the spacecraft, in a synodic reference frame that rotates at the same rate as the orbital motion of the two primary masses. The coordinate frame is centered at the barycenter of the system and oriented such that the x -axis extends from the barycenter toward the smaller primary, the z -axis extends toward the primary bodies' orbit normal, and the y -axis completes the right-handed coordinate frame. In that synodic frame, the two massive bodies are stationary, and the spacecraft moves about the system in non-Keplerian motion [46, 86]. It is convenient to normalize the units in the system such that the following measurements are equal to one: the distance between the two primaries, the sum of the mass of the two primaries, the rotation rate of the system, and the gravitational parameter. The three-body constant, μ , relates all of these normalized measurements and is easily computed by dividing the mass of the smaller primary by the total mass in the system. The equations of motion for the third body in the normalized rotating frame are equal to [86]

$$\ddot{x} = 2\dot{y} + x - (1 - \mu) \frac{x + \mu}{r_1^3} - \mu \frac{x - 1 + \mu}{r_2^3} \quad (2.1)$$

$$\ddot{y} = -2\dot{x} + y - (1 - \mu) \frac{y}{r_1^3} - \mu \frac{y}{r_2^3} \quad (2.2)$$

$$\ddot{z} = -(1 - \mu) \frac{z}{r_1^3} - \mu \frac{z}{r_2^3} \quad (2.3)$$

where r_1 and r_2 are equal to the distance from the third body to the larger and smaller primary, respectively

$$r_1^2 = (x + \mu)^2 + y^2 + z^2 \quad (2.4)$$

$$r_2^2 = (x - 1 + \mu)^2 + y^2 + z^2 \quad (2.5)$$

The dynamics in the circular restricted three-body system depend only on the three-body constant, μ . Furthermore, as μ goes to zero, the dynamics approach two-body dynamics, although represented in a rotating frame.

2.5.1.2 Lagrange Points There are five well-known equilibrium solutions to the CRTBP, known as the five Lagrange points [86], or the five libration points. These points are referred to as L_1 – L_5 ; this book adopts the nomenclature that L_1 lies between the two primary masses and L_2 lies on the far side of the smaller primary, relative to the barycenter of the system. The Lagrange points in the Earth–Moon system are abbreviated using the nomenclature LL_1 – LL_5 ; the Sun–Earth Lagrange points are abbreviated EL_1 – EL_5 . The seven Lagrange points near the Earth are depicted in Fig. 2-2. More discussion about their locations and dynamics are provided in Sections 2.6.2 and 2.6.10.

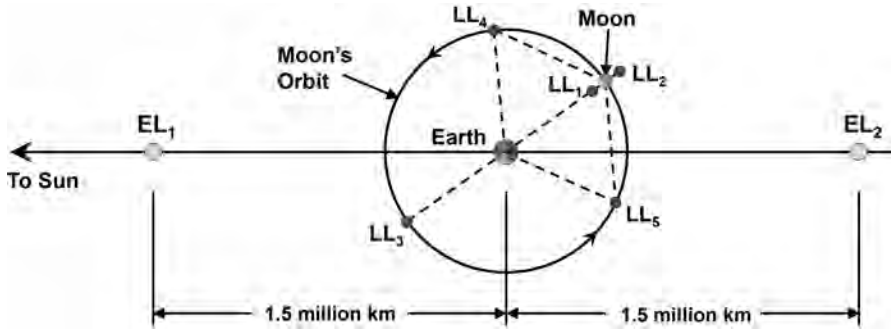


Figure 2-2 A plot depicting the relative proximity of the five Earth–Moon Lagrange points and the two nearby Sun–Earth points (first published in Ref. [97]; reproduced with kind permission from Springer Science+Business Media B.V.).

2.5.1.3 Jacobi Constant The dynamics of the CRTBP permit an integral of motion to exist in the synodic reference frame, known as the Jacobi integral or Jacobi constant [46, 97, 98]. The Jacobi constant of a spacecraft in the three-body system may be written simply as

$$C = 2U - V^2, \quad \text{where} \quad (2.6)$$

$$U = \frac{1}{2}(x^2 + y^2) + \frac{1 - \mu}{r_1} + \frac{\mu}{r_2} \quad (2.7)$$

$$V^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2 \quad (2.8)$$

The spacecraft's position and velocity coordinates in Equations 2.1–2.8 are given in nondimensional normalized synodic coordinates, relative to the barycenter of the three-body system. The Jacobi constant of a spacecraft moving in the CRTBP may not change unless the spacecraft is perturbed in some way other than by the gravitational attraction of the two primary bodies.

It is useful to consider the Jacobi constant of spacecraft in different practical orbits in order to place the value of the Jacobi constant in context. From two-body analyses, we know that spacecraft in orbits about the Earth below the geosynchronous Earth orbit (GEO) belt are only slightly perturbed by the gravity of the Moon. A spacecraft in a 185-km near-circular orbit about the Earth has a Jacobi constant of approximately 58.0, though there is some variation depending on the location of the Moon, the time of year, and the inclination of the orbit. If the spacecraft's orbital altitude is increased, its Jacobi constant decreases. A spacecraft in a 1000-km near-circular orbit has a Jacobi constant near 51.5, a GPS satellite has a Jacobi constant near 14.6, a GEO satellite has a Jacobi constant near 9.6, and so forth. The same trend exists for orbits about the Moon. A spacecraft in a near-circular lunar orbit at an altitude near 100 km has a Jacobi constant near 5.5, and a satellite in a lunar orbit at an altitude near 1000 km has a Jacobi constant near 4.7, and so forth. A spacecraft on a direct transfer to the Moon has a Jacobi constant in the vicinity of 2.3, depending on the particulars

of the transfer. Likewise, a spacecraft on a low-energy transfer to the Moon departs with a Jacobi constant of about 0.8, though it changes significantly before it arrives at the Moon due to the gravity of the Sun.

One observes that a spacecraft with a smaller Jacobi constant can traverse further from either central body. A useful analysis is to identify the boundary of possible motion for a spacecraft with a particular Jacobi constant. These boundaries are computed by setting the velocity of the spacecraft equal to zero in Eq. (2.6); they are hence known as zero-velocity curves. Figure 2-3 illustrates the zero-velocity curves for several Jacobi constants for motion in the x - y plane in the Earth–Moon system.

2.5.1.4 Forbidden Regions A spacecraft traversing the Earth–Moon system with a Jacobi constant less than 2.988 (the approximate Jacobi constant of the L_4 and L_5 points) can theoretically reach any point in the entire system. Its velocity in the rotating frame will decrease to a minimum if it traverses through the L_4 or L_5 points, but no region is inaccessible. Any spacecraft that has a Jacobi constant greater than about 2.988 cannot physically reach all regions, but is bounded by the zero-velocity curves to regions of space that permit its Jacobi constant value. Those regions in space that the spacecraft cannot reach are known as *forbidden regions*.

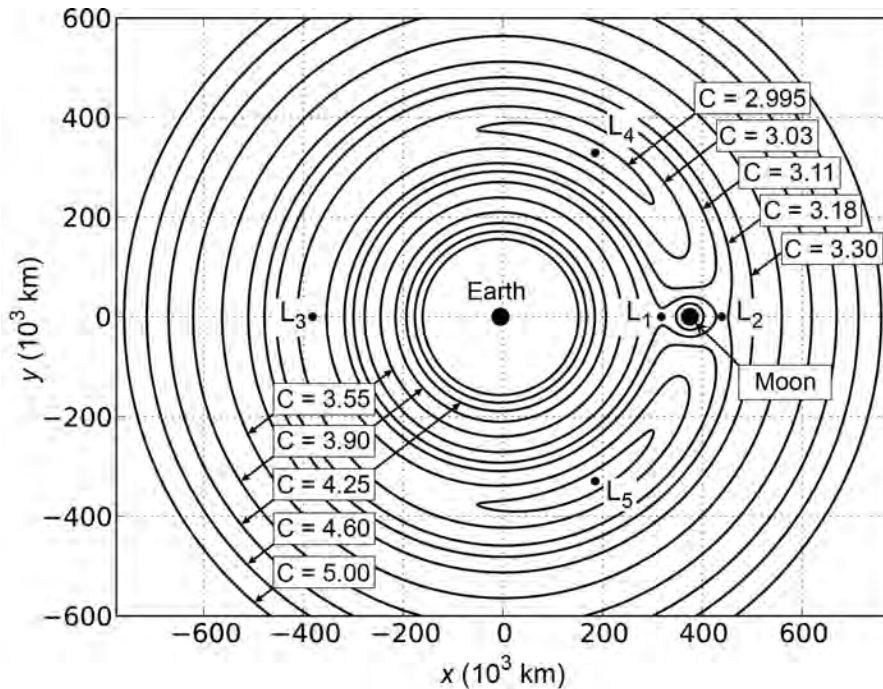


Figure 2-3 An illustration of zero-velocity curves for several Jacobi constant values in the planar Earth–Moon system.

Consider a spacecraft with a Jacobi constant of 3.18 in the Earth–Moon system (see Fig. 2-3). Its forbidden region encircles the Earth–Moon system, including the L_2 – L_5 points. If the spacecraft begins at a point near the Earth or Moon, it can traverse anywhere between the Earth and Moon within the corresponding zero-velocity curve, including transferring through the gap at the L_1 point. If the spacecraft begins well outside of the system, then it must remain beyond the zero-velocity curve. It cannot match the angular velocity of the rotating frame any nearer than its zero-velocity curve.

2.5.1.5 Symmetries The existence of symmetries in the CRTBP is of particular interest for some of the analyses encountered in this book. One symmetry that is quite useful was demonstrated by Miele in his examination of image trajectories in the Earth–Moon space [99]. He showed that if $(x, y, z, \dot{x}, \dot{y}, \dot{z}, t)$ is a solution in the CRTBP, then $(x, -y, z, -\dot{x}, \dot{y}, -\dot{z}, -t)$ is also a solution. In other words, if a trajectory is reflected about the xz plane, a valid trajectory is obtained by traveling along the reflected trajectory in reverse. This property eliminates the need to compute approach and departure trajectories separately in the CRTBP. Another useful symmetry that exists in the CRTBP is that if $(x, y, z, \dot{x}, \dot{y}, \dot{z}, t)$ is a solution in the CRTBP, then $(x, y, -z, \dot{x}, \dot{y}, -\dot{z}, t)$ is also a solution. This permits trajectories to have a *Northern* and a *Southern* variety. Since the CRTBP approximates many aspects of the real Solar System, one may also frequently use CRTBP reflections as approximations for trajectories in the real Solar System.

2.5.2 Patched Three-Body Model

The patched three-body model [38–40, 45, 46] uses purely three-body dynamics to model the motion of a spacecraft in the presence of the Sun, Earth, and Moon. It retains many of the desirable characteristics of the CRTBP, while permitting a spacecraft in the near-Earth environment to be affected by all three massive bodies, albeit only two massive bodies at any given moment. When the spacecraft is near the Moon, the spacecraft’s motion is modeled by the Earth–Moon three-body system. Otherwise, the spacecraft’s motion is modeled by the Sun–Earth three-body system, where the secondary body is the barycenter of the Earth and Moon. For simplicity it is assumed that the Earth–Moon system is coplanar with the Sun–Earth system. The boundary of these two systems is referred to as the three-body sphere of influence (3BSOI); it is analogous to the two-body sphere of influence used in the patched conic method of interplanetary mission design.

Parker describes the 3BSOI as the boundary of a sphere centered at the Moon with a radius r_{SOI} computed using the following relationship [46]

$$r_{\text{SOI}} = a \left(\frac{m_{\text{Moon}}}{m_{\text{Sun}}} \right)^{2/5} \quad (2.9)$$

where m_{Moon} and m_{Sun} are the masses of the Moon and Sun, respectively, and a is the average distance between the Sun and Moon, equal to approximately 1 astronomical

unit (AU). Thus, the 3BSOI has a radius of approximately 159,200 km, which is large enough to include LL_1 and LL_2 .

2.5.3 JPL Ephemeris

The Jet Propulsion Laboratory and the California Institute of Technology have developed the DE421 Planetary and Lunar Ephemerides, which is the most accurate model of the Solar System used in this work. The model includes ephemerides of the positions and velocities of the Sun, the four terrestrial planets, the four gas-giant planets, the Pluto/Charon system, and the Moon [100]. The lunar orbit is accurate to within a meter; the orbits of Earth, Mars, and Venus are accurate to within a kilometer [100].

Low-energy lunar transfers modeled in the patched three-body model repeat perfectly from one synodic month to the next, since the dynamics and the Sun–Earth–Moon geometry are perfectly symmetric. It is often possible to build a very similar low-energy lunar transfer from one month to the next in the DE421 model of the Solar System, but its characteristics will vary in each month. This will be further discussed in later chapters.

2.5.3.1 Earth Orbit The Earth–Moon system orbits the Sun in a nearly circular orbit, but its nonzero eccentricity has an impact on the performance of a particular low-energy lunar transfer from one month to the next. Furthermore, its orbit changes over time due to the influence of Jupiter and the other planets. Figure 2-4 illustrates the Earth’s osculating eccentricity, semi-major axis, and inclination over time in the DE421 model of the Solar System, relative to the Sun. One notices a nearly annual signal in the eccentricity and a bi-annual signal in the semi-major axis. This is predominantly due to the influence of Jupiter’s gravity, which has a synodic period of about 399 days.

2.5.3.2 Lunar Orbit For the purposes of mission planning, the Moon’s orbit about the Earth may be modeled as circular and coplanar with Earth’s orbit about the Sun. In reality, the Moon’s orbit about the Earth is inclined by about 5.15 deg relative to the ecliptic, and it has an average eccentricity of about 0.05490—quite a bit higher than the Earth’s orbital eccentricity. Figure 2-5 illustrates the Moon’s osculating eccentricity, semi-major axis, and inclination over time in the DE421 model of the Solar System, relative to the Earth. The Moon’s orbit is strongly perturbed by the gravity of the Sun on several time scales. First, one can see a very clear signal in the time series of the Moon’s orbital parameters that has a frequency of about 29.53 days, corresponding to the length of an average synodic month. Another very strong signal in the time series of the Moon’s orbital parameters has a frequency of about 6 months, corresponding to the bi-annual impact of the Earth’s orbit about the Sun. The relative orientation of the Moon’s orbit to the Sun cycles over the course of a year, as well as the distance to the Sun. Both the orientation and the distance have a direct effect on the orbit. In addition to the solar perturbation, the planets Venus and Jupiter impact the lunar orbit, as does the Earth’s asymmetric gravity field.

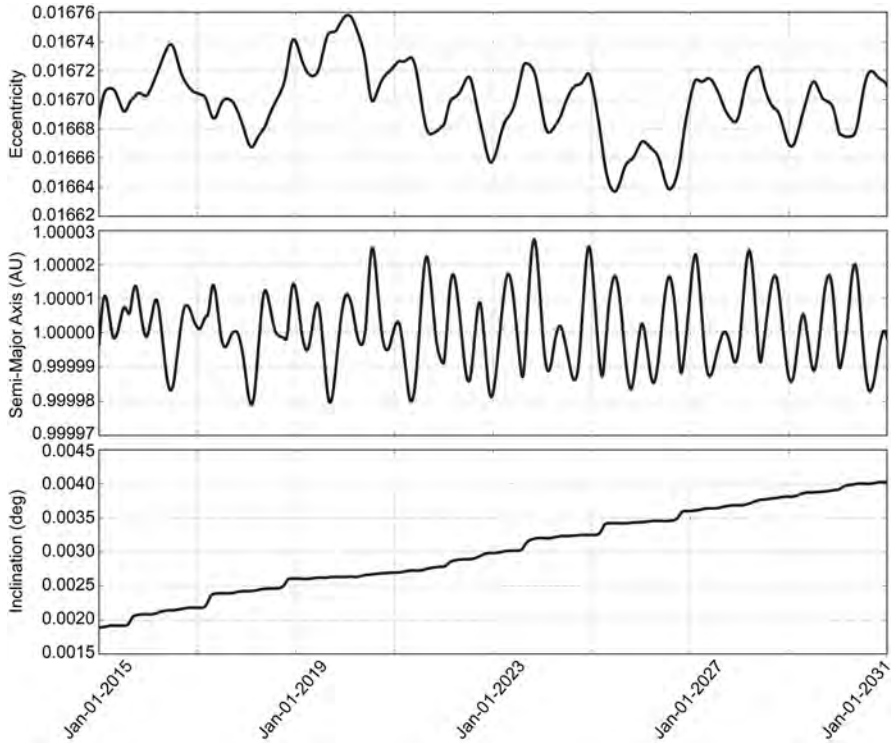


Figure 2-4 The instantaneous eccentricity (top), semi-major axis (middle), and inclination (bottom) of the Earth–Moon barycenter over time relative to the Sun in the EMO2000 coordinate frame.

In addition to the three orbital parameters illustrated in Fig. 2-5, the orientation of the Moon’s orbit about the Earth undergoes both secular and periodic variations. Most notably, the Moon’s orbit precesses about the ecliptic North Pole. The period of regression of the longitude of the lunar orbit’s ascending node (Ω) is equal to about 18.6 years. The period of precession of the lunar orbit’s argument of periaapse (ω) is equal to about 6.0 years. Finally, the period of precession of the longitude of periaapse ($\Omega + \omega$) is equal to about 8.85 years.

2.6 LOW-ENERGY MISSION DESIGN

The field of low-energy mission design relates to the study of trajectories that traverse unstable three-body orbits and take advantage of the dynamics to perform orbit transfers using very little fuel. This section will describe three-body orbits, their unstable manifolds, and how to construct low-energy transfers between them. Indeed, an ex-

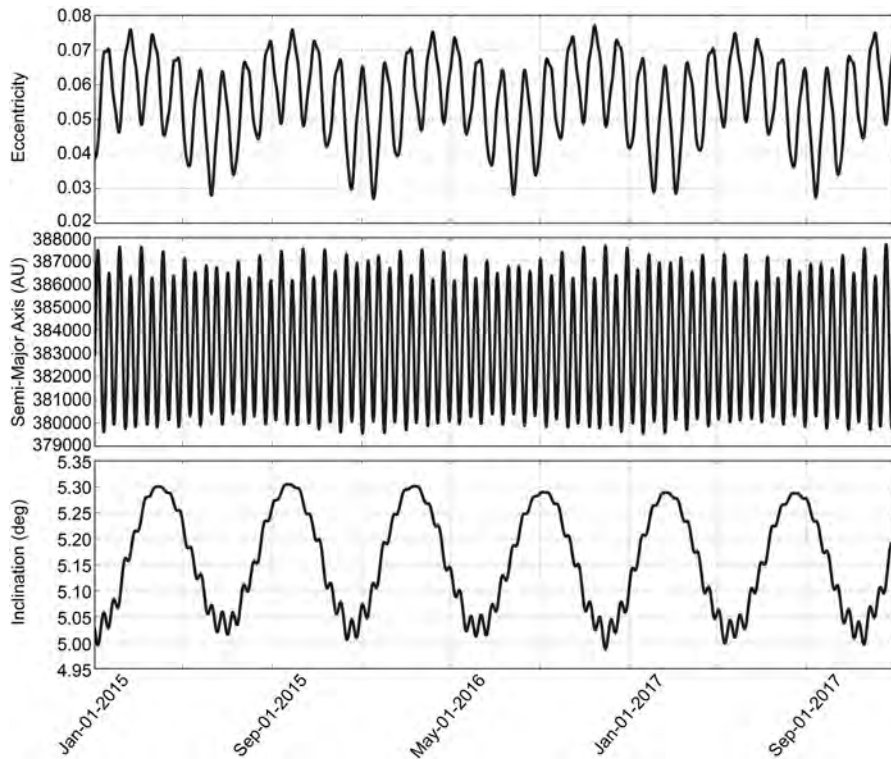


Figure 2-5 The instantaneous eccentricity (top), semi-major axis (middle), and inclination (bottom) of the Moon over time relative to the Earth in the EMO2000 coordinate frame.

ample low-energy lunar transfer is described later in dynamical systems terminology as a trajectory that first departs the Earth on the stable manifold of a Sun–Earth orbit, transfers from the stable manifold to an unstable manifold, and traverses that until it intersects the stable manifold of an orbit in the Earth–Moon system. This section describes dynamical systems analyses and how those methods may be applied to practical spacecraft mission design.

2.6.1 Dynamical Systems Theory

A dynamical system may be described as a state space with a set of rules, where the rules govern the evolution of objects’ states through time within the system. The rules are deterministic; that is, the evolution of a state through a particular amount of time yields only one future state.

There are different types of dynamical systems depending on the mathematics involved and the allowable values of time. If time is continuous, capable of taking

any value in the set of real numbers, then the dynamical system is smooth and is called a *flow*. If time may only take discrete values, then the dynamical system is a *map*. Models of the Solar System are generally described by flows. A spacecraft's trajectory in such dynamical systems is the set of states that the spacecraft will take as it moves through time, given its initial state. When integrating the equations of motion for a spacecraft through time using a machine, time cannot truly take on any value in the set of real numbers. The process of integration is a mapping of the spacecraft's state from one point in the state space to another point. A spacecraft's mapped trajectory is therefore only an approximation of the true trajectory.

There are many techniques that are commonly used to analyze dynamical systems. In this work, we begin our analysis of the CRTBP by identifying fixed points and periodic orbits that exist in the system. We continue by studying the stability of those solutions. These techniques provide an understanding of the motion of trajectories near those solutions. Further analysis gradually provides more information about the motion of trajectories throughout the dynamical system.

2.6.2 Solutions to the CRTBP

The CRTBP is a good example of a system in which dynamical systems methods of analysis work well. The CRTBP contains five fixed-point solutions and an infinite number of periodic orbit solutions. The characterization of these solutions helps to understand the flow of particles and spacecraft in the system. Useful trajectories may then be constructed that take advantage of the flow in the system, rather than forcing their way through the system. The following sections describe some of the simplest solutions to the CRTBP.

2.6.2.1 Fixed-Point Solutions: Five Lagrange Points The most basic solutions to the CRTBP are fixed-point solutions, that is, the trajectories in the CRTBP that particles may follow such that they stay at rest in the system indefinitely. There are five such fixed-point solutions in the CRTBP, namely, the five Lagrange points. These points were introduced in Section 2.5.1 and are displayed again in Fig. 2-6 for the case of the Earth–Moon CRTBP.

The locations of the five Lagrange points in the Sun–Earth and Earth–Moon circular three-body systems are given in Table 2-3, using the planetary masses and distances provided in the Constants, page 382. Appendix A provides an analytical derivation for the locations of the five Lagrange points for any three-body system, as well as algorithms to determine their locations. Table 2-4 summarizes the Jacobi constant of each of the five Lagrange points for both three-body systems.

2.6.2.2 Periodic and Quasiperiodic Orbit Solutions The CRTBP permits the existence of numerous families of periodic and quasiperiodic orbits. Authors have been studying such orbits since the late 1800s, though the introduction of modern computing capability dramatically improved the quantity and complexity of orbits that could be generated.

A periodic orbit in the three-body system may just be a two-body orbit about one of the bodies that is slightly perturbed by the other massive body and is in resonance

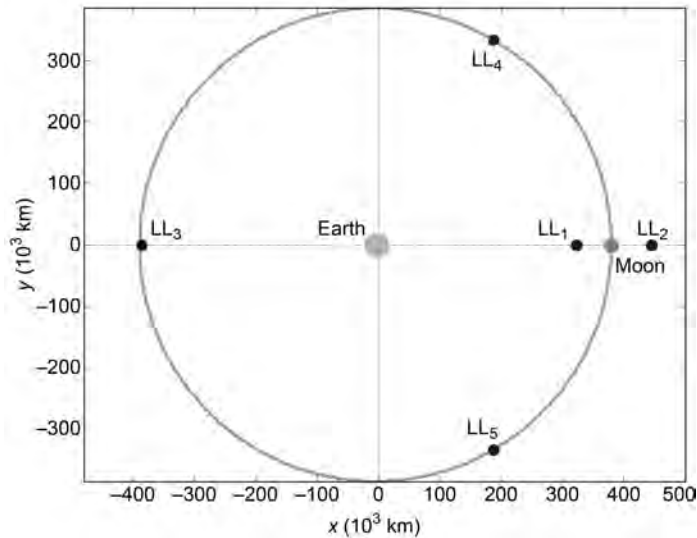


Figure 2-6 The locations of the five Lagrange points in the Earth–Moon CRTBP.

Table 2-3 The locations of the five Lagrange points in the Sun–Earth and Earth–Moon circular three-body systems. The positions are given in nondimensional normalized units and kilometers with respect to the barycenter of the system, assuming the masses and distances given in the Methodology Introduction in Section 2.1.

Lagrange Point	Position in normalized units			Position in kilometers		
	x	y	z	x	y	z
Sun–Earth	L ₁	0.9899859823	0	0	148,099,795.0	0
	L ₂	1.0100752000	0	0	151,105,099.2	0
	L ₃	−1.0000012670	0	0	−149,598,060.2	0
	L ₄	0.4999969596	0.8660254038	0	74,798,480.5	129,555,556.4
	L ₅	0.4999969596	−0.8660254038	0	74,798,480.5	−129,555,556.4
Earth–Moon	L ₁	0.8369151324	0	0	321,710.177	0
	L ₂	1.1556821603	0	0	444,244.222	0
	L ₃	−1.0050626453	0	0	−386,346.081	0
	L ₄	0.4878494157	0.8660254038	0	187,529.315	332,900.165
	L ₅	0.4878494157	−0.8660254038	0	187,529.315	−332,900.165

with the motion of the primaries, that is, a low Earth orbit with a period that is perfectly resonant with a sidereal month. Such an orbit has characteristics not unlike any other low Earth orbit, except that its orbital elements were carefully selected to be periodic with the Moon in the presence of the Moon’s perturbations. Further, such a low Earth orbit is not quite periodic from one revolution to the next about the Earth,

Table 2-4 The Jacobi constant of each Lagrange point in the Earth–Moon and the Sun–Earth three-body systems, given in normalized coordinates.

Lagrange Point	Earth–Moon C	Sun–Earth C
L_1	3.18834129	3.00089794
L_2	3.17216060	3.00089388
L_3	3.01214717	3.00000304
L_4	2.98799703	2.99999696
L_5	2.98799703	2.99999696

due to the Moon’s perturbations; it is only perfectly periodic over the course of its resonant cycle with the Moon.

Alternatively, one can construct a trajectory that carefully balances the three-body dynamics and can only exist in any form under the gravitational attraction of both bodies. Examples of three families of such periodic orbits are illustrated in Fig. 2-7. These orbits include *libration orbits* about the Earth–Moon L_1 and L_2 points and *distant prograde orbits* about the Moon. It should be noted that the smallest distant prograde orbits are very similar to two-body orbits about the Moon in resonance with the orbital motion of the Earth and Moon. Clearly the libration orbits about the Lagrange points only exist within a three-body system.

Between the 1890s and the 1930s, George Darwin [102, 103], George Hill [104], Henry Plummer [105], Forest Moulton [106], Elis Strömberg [107], and their col-

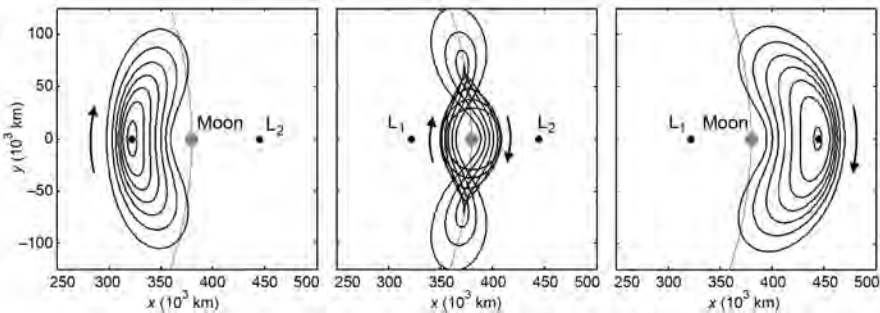


Figure 2-7 Several example orbits from three families of unstable periodic Earth–Moon three-body orbits, viewed from above in the Earth–Moon synodic reference frame. The orbits shown are from the family of Lyapunov orbits about L_1 (left), the family of distant prograde orbits about the Moon (center), and the family of Lyapunov orbits about L_2 (right). The arrows indicate the motion of objects traversing these orbits; the Moon’s orbital radius about the Earth–Moon barycenter is shown in gray for reference [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

leagues contributed to the discovery of the first known periodic orbits in the circular restricted three-body problem. Over the course of 40 years, it is unlikely that more than 150 orbits were ever computed [108]; however, the general aspects of orbits in the three-body problem became well-understood.

In the 1960s, modern computers became accessible, and numerical techniques could be used to swiftly identify and compute periodic orbits. In 1968, Roger Broucke published a large catalog of families of planar periodic orbits that exist in the CRTBP with Earth–Moon masses [108]. Also in the 1960s, researchers computed and cataloged a large number of three-dimensional periodic orbits; significant contributors include Michel Hénon [109–113], Arenstorf [114], Goudas [115], Bray and Goudas [116, 117], and Kolenkiewicz and Carpenter [118], among others. Halo and quasi-halo orbits were discovered and analyzed beginning in the late 1960s (see, for example, Farquhar [119], Farquhar and Kamel [120], Breakwell and Brown [121], and Howell [122]). In 1980, David Richardson used the Lindstedt–Poincaré method to analytically produce periodic orbits, such as halo orbits, about the collinear libration points [123]. Additional work was accomplished toward the end of the 20th Century studying Lissajous and other quasi-halo orbits (see, for example, Farquhar and Kamel [120], Howell and Pernicka [124], and Gómez et al. [67, 125]). Many authors have studied how to take advantage of libration orbits for practical spacecraft missions, including scientific missions such as *WMAP* and *SOHO*, communication relays [5–7, 11], transportation nodes [14, 126], and navigation services [8, 10–13, 127, 128].

In this section, we demonstrate how to analytically construct one set of periodic and quasiperiodic orbits that exist about each of the collinear Lagrange points. This demonstration sheds light on why many periodic orbits exist [106, 123, 124].

We begin by translating the origin of the synodic frame to one of the collinear libration points, L_i . The parameter γ is defined to be equal to the distance from L_i to the smaller primary. The value of γ is positive when referring to L_2 and negative when referring to L_1 and L_3 . The new position coordinates x' , y' , and z' are thus defined by the following

$$\begin{aligned} x' &= x - (1 - \mu + \gamma) \\ y' &= y \\ z' &= z \end{aligned}$$

If we now linearize the equations of motion of the CRTBP given in Eqs. (2.1)–(2.3) under this transformation, we find the following

$$\begin{aligned} \ddot{x}' - 2\dot{y}' - (1 + 2c)x' &= 0 \\ \ddot{y}' + 2\dot{x}' + (c - 1)y' &= 0 \\ \ddot{z}' + cz' &= 0 \end{aligned} \tag{2.10}$$

where c is a constant coefficient. The analytical solution to the out-of-plane z motion describes simple harmonic motion. The solution for the in-plane x – y motion involves a characteristic equation that has two real roots and two imaginary roots. The roots

represent modes of motion, one divergent and one nondivergent. If the nondivergent mode is excited, then the solution is bounded and may be written as

$$\begin{aligned}x' &= -kA_y \cos(\lambda t + \phi) \\y' &= A_y \sin(\lambda t + \phi) \\z' &= A_z \sin(\nu t + \psi)\end{aligned}\tag{2.11}$$

This motion is described by six variables: the amplitudes of the in-plane and out-of-plane motion (A_y and A_z), the frequency of oscillation in the in-plane and out-of-plane motion (λ and ν), and the phase angles for the in-plane and out-of-plane motion (ϕ and ψ). The linearized approximation to the analytical solution for periodic motion about a Lagrange point may thus be characterized by oscillatory motion. If the two frequencies λ and ν are equal or otherwise commensurate, the resulting motion will be periodic; if the frequencies are incommensurate, the resulting motion will be quasiperiodic. The periodic orbits whose frequencies are *equal* are known as halo orbits, the more general quasiperiodic trajectories are known as Lissajous orbits or quasi-halo orbits. A portion of the family of halo orbits about the Earth–Moon L_2 is shown in Fig. 2-8, and characteristic views of several types of Lissajous orbits are shown in Fig. 2-9.

It should also be noted that there is a symmetry that exists in the CRTBP, as described in Section 2.5.1. If the CRTBP permits an orbit to exist, then it also permits a symmetric orbit to exist that is a reflection across the $y = 0$ plane. Hence, there are two families of halo orbits, a northern and a southern. By convention, if a spacecraft spends more than half of its time above the $y = 0$ plane in a halo orbit, then the spacecraft is following a northern halo orbit.

If A_z is set to zero in Eq. (2.11) the resulting orbits are planar and are known as Lyapunov orbits. Figure 2-7 shows a portion of the families of Lyapunov orbits about L_1 and L_2 in the Earth–Moon system.

These orbits may be constructed analytically since the linearization process near one of the Lagrange points produces a good approximation of the true dynamics found in the system. Other orbits do not have well-described linear approximations and must be constructed numerically. The process of numerically constructing simple periodic orbits is discussed in Section 2.6.6.

Periodic orbits in the three-body system exist that revolve about all five Lagrange points, the primary, the secondary, and also about the entire system. Periodic orbits exist that revolve about either body in a prograde sense and a retrograde sense. One may construct simple symmetric periodic orbits, such as those illustrated in this section, and one may construct asymmetric, complex orbits. A wide variety of periodic orbits exist that may be useful to the mission planner.

2.6.2.3 Orbit Parameters An orbit and a position in that orbit may be uniquely specified in the two-body problem using six parameters. Typical sets of two-body parameters include the Cartesian and Keplerian sets. Parameterization of orbits in the three-body problem has proven to be much more difficult, since there are no general analytical solutions to the system. Dynamical systems theory is very useful

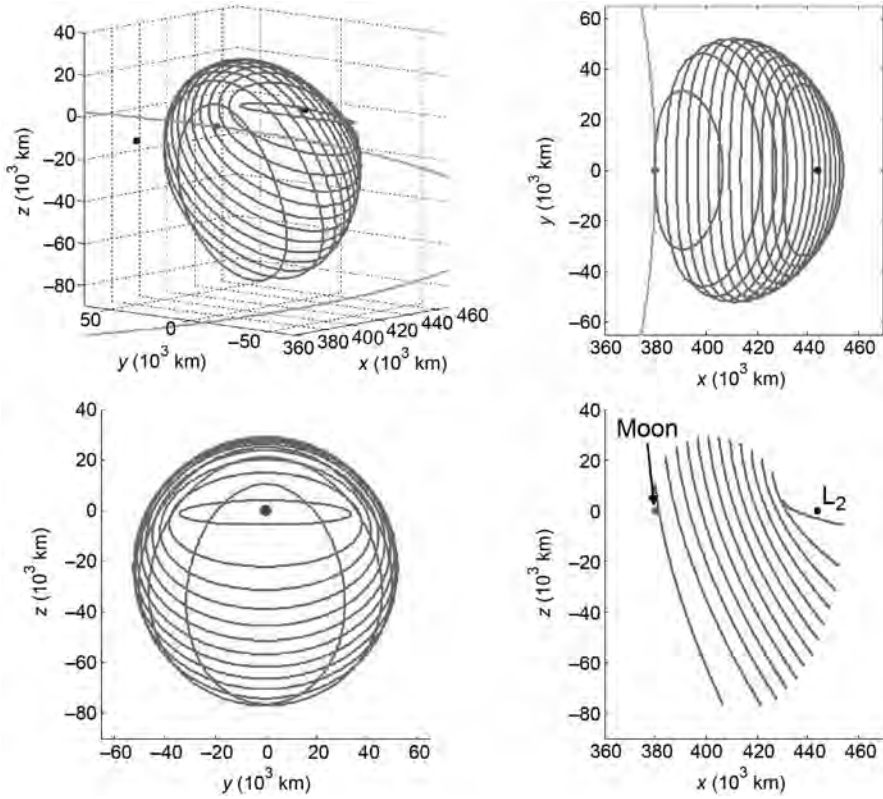


Figure 2-8 A portion of the family of halo orbits about L_2 in the Earth–Moon system, shown from four perspectives.

in this regard because the methodology lends itself to many useful parameters. One such parameter, τ , is useful when describing periodic orbit solutions to the CRTBP. This parameter is described here; others are introduced in later chapters as their uses become apparent.

The parameter τ mimics the two-body mean anomaly. For the case of halo orbits, and other symmetric periodic orbits in the CRTBP, τ advances at a steady rate over time, beginning at one landmark (typically where the orbit pierces the $y = 0$ plane) and resetting when it completes an entire period. In some studies, τ takes on values in the range of 0–360 deg, much like the mean anomaly [11]. In other studies, τ is defined to take on values in the range of 0–1, indicating the periodic revolution number [46]. Most libration orbits, for example, halo and Lyapunov orbits, have a shape that resembles a conic section; in those cases it is intuitive to use an angular unit of measurement for τ . However, there are many classes of periodic orbits that

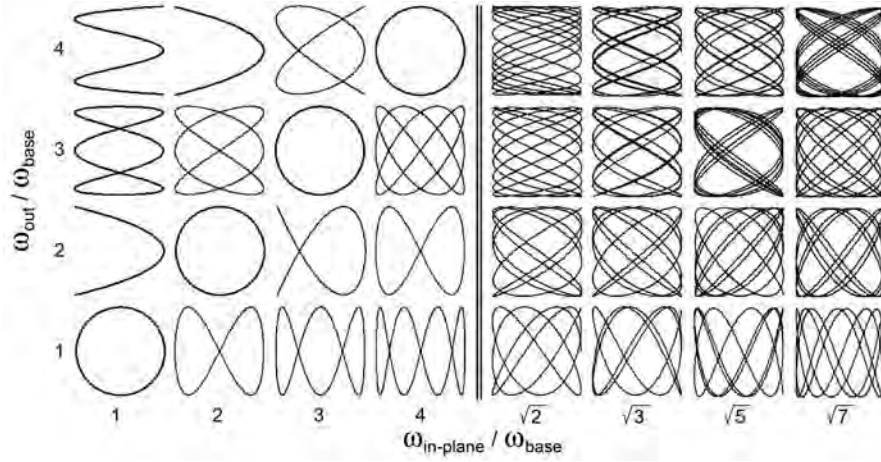


Figure 2-9 A sample of Lissajous curves representing the view of Lissajous orbits in the Earth–Moon system as viewed from an observer at the Earth looking toward the Moon; $\omega_{\text{in-plane}}$ and ω_{out} are multiples of some base frequency ω_{base} . The curves on the left are perfectly periodic; the curves on the right have incommensurate frequencies and have only been propagated for a short amount of time (first published in Ref. [97]; reproduced with kind permission from Springer Science+Business Media B.V.).

do not resemble any sort of conic section, and it may be confusing to refer to τ in angular units. Figure 2-10 shows two orbits, demonstrating how τ advances along each orbit, where τ has been represented as a revolution number for a complex orbit and as an angle for an L_2 libration orbit.

2.6.3 Poincaré Maps

A Poincaré map is a useful tool for analyzing dynamical systems and is often used to identify orbits and orbit transfers in a complex system. A Poincaré map is created by intersecting a trajectory in the n -dimensional flow $\dot{\mathbf{x}} = f(\mathbf{x})$ by an $(n-1)$ -dimensional surface of section Σ . Thus, the Poincaré mapping replaces the flow of an n^{th} order system with a discrete system of order $(n-1)$ [129]. A Poincaré mapping, P , may be described as a function that maps the state of a trajectory at the k^{th} intersection with the surface of section, \mathbf{x}_k , to the next intersection, \mathbf{x}_{k+1}

$$\mathbf{x}_{k+1} = P(\mathbf{x}_k) \quad (2.12)$$

If a trajectory pierces Σ at the state \mathbf{x}^* at time t and then returns to \mathbf{x}^* at time $t + T$, then one may conclude that the trajectory is a periodic orbit with a period T [130].

There are three different types of Poincaré maps considered in this research, defined as follows [130]:

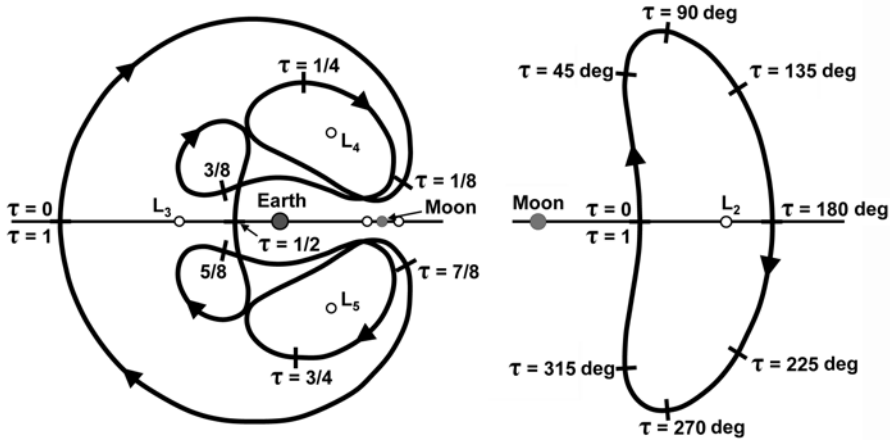


Figure 2-10 The two orbits shown demonstrate how the parameter τ advances from 0 to 1 about a complex orbit (left) or from 0 deg to 360 deg about a libration orbit (right). Both orbits are viewed from above in the Earth–Moon CRTBP synodic frame.

- P_+ : The Poincaré map created from only the positive intersections of the trajectory with the surface of section. For instance, in the CRTBP, Σ may be defined as a y - z plane set to some x -value and P_+ includes only those intersections that have positive values of \dot{x} .
- P_- : The Poincaré map created from only the negative intersections of the trajectory with the surface of section.
- P_{\pm} : The Poincaré map created from all intersections of the trajectory with the surface of section.

The maps P_+ and P_- are called one-sided maps, while P_{\pm} is called a two-sided map [130]. Figure 2-11 provides a simple illustration of a one-sided Poincaré mapping of two orbits, where one is periodic and one is not immediately periodic.

2.6.4 The State Transition and Monodromy Matrices

The state transition matrix $\Phi(t, t_0)$ monitors the divergent dynamics along a trajectory. Essentially, it approximates how a slight deviation in any of the state variables propagates along the trajectory. Its practical uses are twofold in this study:

1. to provide a means of adjusting the initial conditions of a trajectory to correct for unwanted motion, and
2. to provide information about the stability of an orbit, including the orientation of the eigenvectors along the orbit.

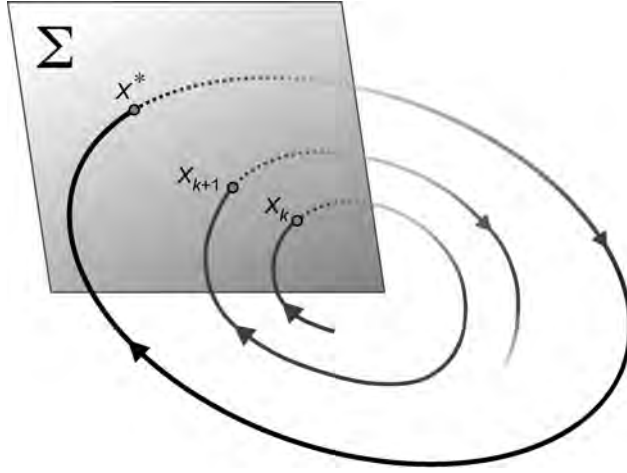


Figure 2-11 An illustration of a one-sided Poincaré mapping of two trajectories. The point x^* indicates a fixed point on the surface, corresponding to a periodic trajectory [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

The second use involves the *monodromy matrix*, a special case of the state transition matrix. We explore (1) in Sections 2.6.5 and 2.6.6 and (2) in Sections 2.6.8 and 2.6.10. This section discusses how to construct the state transition matrix and the monodromy matrix.

Let us define the state vector \mathbf{X} to be a column vector that contains all of the state variables of interest in the system. In most cases, one is usually only interested in computing the six state variables of a particle or spacecraft in a system. Hence, \mathbf{X} is defined as

$$\mathbf{X} = [x \ y \ z \ \dot{x} \ \dot{y} \ \dot{z}]^T$$

Then the state transition matrix is a 6×6 matrix composed of the partial derivatives of the state

$$\Phi(t, t_0) = \frac{\partial \mathbf{X}(t)}{\partial \mathbf{X}(t_0)} \quad (2.13)$$

with initial conditions $\Phi(t_0, t_0) = I$. The state transition matrix is propagated using the following relationship

$$\dot{\Phi}(t, t_0) = A(t)\Phi(t, t_0) \quad (2.14)$$

where the matrix $A(t)$ is equal to

$$A(t) = \frac{\partial \dot{\mathbf{X}}(t)}{\partial \mathbf{X}(t)} \quad (2.15)$$

In the CRTBP, $A(t)$ is equal to

$$A(t) = \begin{bmatrix} 0 & I \\ U_{XX} & 2\Omega \end{bmatrix}, \quad \text{where} \quad \Omega = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.16)$$

and U_{XX} is the symmetric matrix composed of second partial derivatives of U with respect to the third body's position evaluated along the orbit

$$U_{XX} = \begin{bmatrix} \frac{\partial \ddot{x}}{\partial x} & \frac{\partial \ddot{x}}{\partial y} & \frac{\partial \ddot{x}}{\partial z} \\ \frac{\partial \ddot{y}}{\partial x} & \frac{\partial \ddot{y}}{\partial y} & \frac{\partial \ddot{y}}{\partial z} \\ \frac{\partial \ddot{z}}{\partial x} & \frac{\partial \ddot{z}}{\partial y} & \frac{\partial \ddot{z}}{\partial z} \end{bmatrix} \quad (2.17)$$

The monodromy matrix, M , exists for periodic orbits and is computed by propagating the state transition matrix one entire orbit: $M = \Phi(t_0 + P, t_0)$ [131]. After being propagated for a full orbit, the matrix contains information about every region that a spacecraft would pass through along that orbit. This matrix's use is further explored in Sections 2.6.8 and 2.6.10.

2.6.5 Differential Correction

Differential correction, as it is implemented here, is a process by which the state transition matrix is used to change a set of initial conditions in order to better satisfy a set of criteria. It is a targeting scheme that converges on its constraints very swiftly within the basin of convergence. Two types of differential correction routines are used in this work: single-shooting and multiple-shooting correction routines.

2.6.5.1 Single-Shooting Differential Correction In the standard single-shooting differential correction routine used in this work, a spacecraft begins at some state \mathbf{X}_0 , following a nominal trajectory $T(t)$, where \mathbf{X}_0 is composed of a position vector \mathbf{R}_0 and a velocity vector \mathbf{V}_0 . It is desired that the spacecraft's trajectory be shifted such that at a later time, t_f , the trajectory encounters a desired state $\hat{\mathbf{X}}_f$ (including a desired position vector $\hat{\mathbf{R}}_f$ and a velocity vector $\hat{\mathbf{V}}_f$). There are usually two constraints to the problem: (1) the spacecraft's initial position may not change; and (2) the spacecraft's new trajectory leads it to a final specified position vector $\hat{\mathbf{R}}_f$. The routine is allowed to vary the initial velocity of the spacecraft (simulating a change in velocity (ΔV) in the mission design), and is oftentimes allowed to vary the time at which the spacecraft arrives at its final desired position. The velocity of the spacecraft at the final position is usually a free variable, and mission designers typically plan to perform an additional ΔV at that time. If the routine converges, a new ballistic trajectory is constructed, $\hat{T}(t)$, that satisfies the two conditions

$$\begin{aligned} \hat{T}(\hat{t}_0) &= \hat{\mathbf{X}}_0 && \text{with } \hat{t}_0 = t_0, \hat{\mathbf{R}}_0 = \mathbf{R}_0, \text{ and } \hat{\mathbf{V}}_0 \text{ free} \\ \hat{T}(\hat{t}_f) &= \hat{\mathbf{X}}_f && \text{with } \hat{t}_f \text{ constrained or fixed, } \hat{\mathbf{R}}_f \text{ constrained, and } \hat{\mathbf{V}}_f \text{ free} \end{aligned}$$

This routine is diagrammed in Fig. 2-12.

The single-shooting method uses the state transition matrix $\Phi(t_f, t_0)$ to estimate what change to make in the initial velocity of the state, $\Delta \mathbf{V}_0$, in order to eliminate the deviation in position at the end of the trajectory $\delta \mathbf{R}_f$. The state transition matrix maps perturbations in the state over time using the following linearized equations

$$\begin{aligned} \delta \mathbf{X}_f &= \Phi(t_f, t_0) \Delta \mathbf{X}_0 \quad \text{or} \\ \begin{bmatrix} \delta \mathbf{R}_f \\ \delta \mathbf{V}_f \end{bmatrix} &= \begin{bmatrix} \Phi_{RR}(t_f, t_0) & \Phi_{RV}(t_f, t_0) \\ \Phi_{VR}(t_f, t_0) & \Phi_{VV}(t_f, t_0) \end{bmatrix} \begin{bmatrix} \Delta \mathbf{R}_0 \\ \Delta \mathbf{V}_0 \end{bmatrix} \end{aligned} \quad (2.18)$$

Since $\Delta \mathbf{R}_0 = 0$ and $\delta \mathbf{V}_f$ is unconstrained, we may simplify Eq. (2.18) and solve for $\Delta \mathbf{V}_0$ to find

$$\Delta \mathbf{V}_0 = [\Phi_{RV}(t_f, t_0)]^{-1} \delta \mathbf{R}_f \quad (2.19)$$

Since the state transition matrix is propagated with linearized equations, the algorithm must be iterated until convergence. When the algorithm is converging, each iteration typically improves the solution by a factor of 10, although factors anywhere between 2 and 100 have been observed [46].

2.6.5.2 Multiple-Shooting Differential Correction Multiple-shooting differential correction takes a series of states and adjusts them all simultaneously to construct a complicated trajectory that satisfies a set of constraints. It is very useful when mission designers wish to construct a long trajectory in an unstable environment in the presence of machine precision. For example, the *Genesis* spacecraft departed the Earth, traversed a quasi-halo orbit about the Sun–Earth L_1 point several times, transferred to the vicinity of the Sun–Earth L_2 point, and then returned to Earth. This entire trajectory may in theory be constructed without a single maneuver. However, in this unstable environment, deviations even as small as round-off errors due to machine precision grow exponentially. A computer using finite-precision mathematics

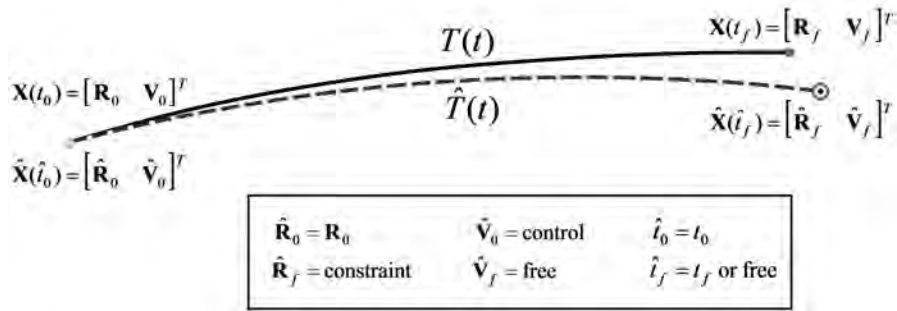


Figure 2-12 The single-shooting differential-correction routine. The solid-line trajectory, $T(t)$, is the initial trajectory; the dashed-line trajectory, $\hat{T}(t)$, is the corrected trajectory that encounters the target position, indicated by a bull's eye, at the target time.

does not normally have the precision required to propagate the spacecraft through its entire trajectory before the round-off errors grow large enough to create a large-scale deviation in the spacecraft's trajectory. To get around this numerical problem, mission designers break the trajectory into many segments and patch the segments together with very small maneuvers. The maneuvers counteract the build-up of propagated round-off errors. They may be as small as desired, depending on the length of each patched segment, and are typically much smaller than any expected station-keeping maneuver. Hence, they are not normally considered to be deterministic maneuvers in the mission.

The multiple-shooting differential correction method is described in detail by Pernicka [132, 133] and by Wilson [134, 135], among other authors [136–138]. This section provides a summary of the process. Section 2.6.5.3 then returns to the details and derives the tools needed to implement the multiple-shooting differential correction technique. The derivation follows that presented by Wilson [134].

The first step in the process of differentially correcting a complex trajectory is to define a series of patchpoints. Several things must be considered when setting up the patchpoints; a discussion of these considerations is given below. From then on, a two-level process is iterated until either the differential corrector fails or each constraint in the system is satisfied. The first level of the process adjusts the patchpoints' velocities; the second level of the process adjusts the patchpoints' positions and epochs. If the patchpoints fall into some basin of convergence for the differential corrector, then the process converges on a continuous trajectory swiftly. The following paragraphs provide more details about the two-level process:

Level 1. The velocities of every patchpoint along the trajectory except the last one are adjusted using the single-shooting differential corrector given in Section 2.6.5.1. The velocities are set such that the position of each segment ends at the following patchpoint. When this step has been completed, the trajectory is continuous, although a spacecraft must perform a ΔV at each patchpoint in order to follow the trajectory.

Level 2. The positions and epochs of every patchpoint, including the last patchpoint, are adjusted using a least-squares method that is designed to reduce the total ΔV cost of the trajectory. The result is a discontinuous trajectory that should require less total ΔV after the following iteration of Level 1.

This iteration process is repeated until the discontinuity at each patchpoint in position and velocity is below some tolerance.

The choice of patchpoints has a strong effect on the differential corrector's performance. First of all, since the single-shooting method is invoked between every adjacent patchpoint, the patchpoints must be close enough to permit the single-shooting method to converge within the desirable tolerance given the numerical precision of the machine. That is, if the patchpoints are too far apart, Level 1 will not converge. Secondly, Level 2 of the two-level process is generally designed with the assumption that each patchpoint is evenly spaced in time. If the patchpoints are not evenly spaced

in time, then the time system should be normalized in some way. This improves the convergence characteristics of the algorithm, but it certainly does not guarantee that the system will converge on a desirable solution. Finally, it has been observed that the differential corrector converges more readily if there are more patchpoints where the dynamics become more unstable, for example, near massive bodies, though the time system often must be normalized for this to benefit the stability of the algorithm.

Section 2.6.5.1 provides the details of Level 1. In order to shed some light onto the functionality of Level 2, let us assume that we have a trajectory built from three patchpoints. The two segments meet in the middle at the second patchpoint, but are not continuous in velocity, that is, there is some velocity deviation $\Delta \mathbf{V}_2$. Let us assume that it is desirable to remove that discontinuity, and to do so we allow the positions and epochs of all three patchpoints to be adjusted. Section 2.6.5.3 provides details about how to numerically compute the variations of the target ($\Delta \mathbf{V}_2$) to the controls ($\mathbf{R}_1, t_1, \mathbf{R}_2, t_2, \mathbf{R}_3, t_3$). In this way, one can construct the following partial derivatives

$$\begin{array}{ccc} \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_1}, & \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_2}, & \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_3}, \\ \frac{\partial \Delta \mathbf{V}_2}{\partial t_1}, & \frac{\partial \Delta \mathbf{V}_2}{\partial t_2}, & \frac{\partial \Delta \mathbf{V}_2}{\partial t_3} \end{array}$$

With these partial derivatives, one can build an approximation of the change in $\Delta \mathbf{V}_2$ when each of the control parameters are perturbed

$$[\delta \Delta \mathbf{V}_2] = \underbrace{\frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_1} \quad \frac{\partial \Delta \mathbf{V}_2}{\partial t_1} \quad \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_2} \quad \frac{\partial \Delta \mathbf{V}_2}{\partial t_2} \quad \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_3} \quad \frac{\partial \Delta \mathbf{V}_2}{\partial t_3}}_{[M]} \begin{bmatrix} \delta \mathbf{R}_1 \\ \delta t_1 \\ \delta \mathbf{R}_2 \\ \delta t_2 \\ \delta \mathbf{R}_3 \\ \delta t_3 \end{bmatrix} \quad (2.20)$$

In general, we wish to determine the appropriate changes to make to each of the control variables in order to reduce the value of $\Delta \mathbf{V}_2$ to zero. The linear system given in Eq. (2.20) is underdetermined; it is common practice to use the smallest Euclidean norm to produce a good solution [134]

$$\begin{bmatrix} \delta \mathbf{R}_1 \\ \delta t_1 \\ \delta \mathbf{R}_2 \\ \delta t_2 \\ \delta \mathbf{R}_3 \\ \delta t_3 \end{bmatrix} = M^T M M^T{}^{-1} [\delta \Delta \mathbf{V}_2] \quad (2.21)$$

These deviations in position and epoch are then added to the patchpoints' states to complete the Level 2 iteration. This example includes only three patchpoints; additional patchpoints may be added on indefinitely. With many patchpoints in the system, the majority of the matrix M is filled with zeros, since each velocity

discontinuity is only dependent on the positions and epochs of the three nearest patchpoints.

In its simplest form, the Level 2 corrections are only constrained by the velocity discontinuities at each patchpoint. Wilson describes how to add many other types of constraints to the differential corrector [134]. Some examples of constraints that may be added include:

- Desirable Position Vector. One may target a particular position vector or position magnitude for any patchpoint in the trajectory. This may be with respect to a point in the coordinate axes or with respect to another body.
- Desirable Inclination. One may target many different orbital parameters, such as the inclination of one or more specified patchpoints.
- Maximum Change in Position. One may limit the differential corrector's capability to change one or more patchpoints' positions during each iteration of Level 2. This helps to keep a trajectory near some initial guess.

Many other types of constraints may be placed on the system. The inclusion of additional constraints is very useful for practical spacecraft missions, where the trajectory must be designed to begin from a particular state or to end at a particular state; however, it does often make it more difficult for the differential corrector to converge.

There are many practical applications of the multiple-shooting differential corrector. To demonstrate its use, we will examine its performance as it is used to differentially correct a periodic halo orbit from the CRTBP into a quasi-halo orbit in the DE421 model of the Solar System. Figure 2-13 provides several representative plots of the differential corrector in action. The plots are exaggerated to demonstrate the procedure clearly. The plot shown in (a) depicts the initial periodic halo orbit in the CRTBP. The trajectory is broken into four segments, separated by five patchpoints as shown in (b), where the fifth patchpoint is coincident with the first in the synodic frame. The first iteration of Level 1, shown in (c), forces the new trajectory to be continuous in position and time in the DE421 model, but permits velocity discontinuities at each interior patchpoint. The five patchpoints' positions and epochs are then adjusted in the first iteration of Level 2 as shown in (d). The plots shown in (e) and (f) give an exaggerated representation of the second iteration of Levels 1 and 2. The plot shown in (g) depicts the trajectory after the third iteration of Level 1; one can see that the trajectory is approaching a continuous trajectory. The plot shown in (h) depicts the final, converged trajectory that is continuous in the DE421 model within some tolerance limits.

Studying Figure 2-13, one can see that the differential corrector permits the ends of the trajectory to be altered substantially since there are no boundary conditions. If more revolutions of the halo orbit were originally sent into the differential corrector, then the final trajectory would resemble the original halo orbit more closely. This process is shown in more detail in Section 2.6.6.3.

The multiple-shooting differential corrector typically operates on a set of patchpoints that define a single trajectory, presumably to be followed by a single spacecraft.

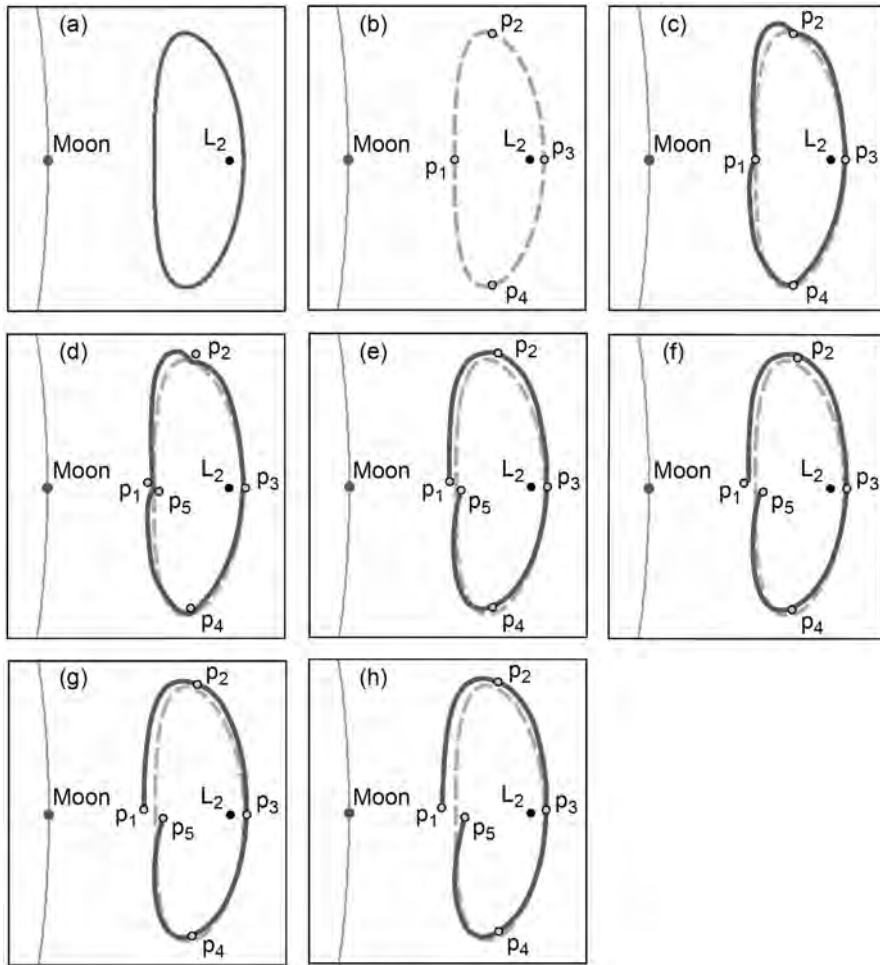


Figure 2-13 An exaggerated demonstration of the implementation of the multiple-shooting differential corrector used to convert a halo orbit from the CRTBP into the DE421 model.

- (a) The initial CRTBP halo orbit.
- (b) The initial placement of patchpoints in the DE421 model.
- (c) Level 1: Differential correction to determine the ΔV s necessary to make the trajectory continuous in the DE421 model.
- (d) Level 2: The adjustments of the patchpoints' positions and epochs to reduce the total ΔV .
- (e) Level 1: The second adjustments of the patchpoints' velocities to make the trajectory continuous.
- (f) Level 2: The second adjustments of the patchpoints' positions and epochs to reduce the total ΔV .
- (g) Level 1: The third adjustments of the patchpoints' velocities to make the trajectory continuous.
- (h) The final converged trajectory in the DE421 model after several additional iterations.

However, the differential corrector may certainly be designed to operate on segments that represent more than one spacecraft, including segments that branch, segments that rendezvous, and/or segments that define a formation.

2.6.5.3 Multiple-Shooting Implementation The multiple-shooting differential corrector is such a useful tool in the design of low-energy trajectories that further attention is given here to derive the algorithms needed to implement it. As described earlier, the multiple-shooting differential corrector involves a process that repeats two steps until a trajectory is generated that satisfies all given constraints. Level 1 is fully described in Section 2.6.5.1, including everything needed to generate software to implement it. Level 2 is introduced in Section 2.6.5.2, but the details have been omitted in order to demonstrate its operation. Those details are provided here.

The engine of the most basic implementation of Level 2 is given by Eq. (2.21), which computes a linear approximation of the changes that must be made to the positions and/or times of the three patchpoints in the scenario in order to reduce the ΔV at the interior patchpoint. The multiple-shooting differential corrector may certainly be extended to include many trajectory segments and a wide variety of constraints. Further, the trajectory segments are not restricted to a single trajectory, but may define multiple trajectories that are simultaneously optimized.

This section begins by describing a basic formulation of Level 2 that involves a single trajectory defined by at least three patchpoints such that the only goal is to reduce the total ΔV required to traverse that trajectory. Next, the section describes the algorithms required to add constraints to the patchpoints in the trajectory. Finally, the section includes a discussion about how to implement the multiple-shooting technique such that it operates on several codependent trajectories simultaneously. In each case, it is always assumed that a ΔV or constraint applied to a particular patchpoint is only affected by the position and/or time of that patchpoint and its neighbors, which is an important feature in the formulation of Level 2.

Basic Level 2. The basic Level 2 formulation is one that operates on a single trajectory and works only to reduce the ΔV of each interior patchpoint. It is assumed that the position and/or time of each patchpoint may be changed to accomplish this goal. Hence, the ΔV at the second patchpoint, ΔV_2 , may be reduced by changing the position, the time, or both of the first, second, and third patchpoints. Any other patchpoints do not directly influence ΔV_2 , though their influences are transmitted through the connecting patchpoints.

Equation (2.21) captures the linear estimate of the change in the positions and times of three patchpoints needed to reduce ΔV_2 , the velocity discontinuity at the interior patchpoint. This expression may be extended to include multiple patchpoints as follows

$$\begin{bmatrix} \delta \mathbf{R}_1 \\ \delta t_1 \\ \delta \mathbf{R}_2 \\ \delta t_2 \\ \delta \mathbf{R}_3 \\ \delta t_3 \\ \vdots \\ \delta \mathbf{R}_n \\ \delta t_n \end{bmatrix} = M^T M M^T{}^{-1} \begin{bmatrix} \delta \Delta \mathbf{V}_2 \\ \delta \Delta \mathbf{V}_3 \\ \vdots \\ \delta \Delta \mathbf{V}_{n-1} \end{bmatrix} \quad (2.22)$$

where the matrix M is constructed using the relationship

$$\begin{bmatrix} \delta \Delta \mathbf{V}_2 \\ \delta \Delta \mathbf{V}_3 \\ \vdots \\ \delta \Delta \mathbf{V}_{n-1} \end{bmatrix} = M \begin{bmatrix} \delta \mathbf{R}_1 \\ \delta t_1 \\ \delta \mathbf{R}_2 \\ \delta t_2 \\ \delta \mathbf{R}_3 \\ \delta t_3 \\ \vdots \\ \delta \mathbf{R}_n \\ \delta t_n \end{bmatrix} \quad (2.23)$$

Thus, M is equal to

$$\begin{bmatrix}
\frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_1} & \frac{\partial \Delta \mathbf{V}_2}{\partial t_1} & \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_2} & \frac{\partial \Delta \mathbf{V}_2}{\partial t_2} & \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_3} & \frac{\partial \Delta \mathbf{V}_2}{\partial t_3} & \mathbf{0} & 0 & \dots & \mathbf{0} & 0 \\
\mathbf{0} & 0 & \frac{\partial \Delta \mathbf{V}_3}{\partial \mathbf{R}_2} & \frac{\partial \Delta \mathbf{V}_3}{\partial t_2} & \frac{\partial \Delta \mathbf{V}_3}{\partial \mathbf{R}_3} & \frac{\partial \Delta \mathbf{V}_3}{\partial t_3} & \frac{\partial \Delta \mathbf{V}_3}{\partial \mathbf{R}_4} & \frac{\partial \Delta \mathbf{V}_3}{\partial t_4} & \dots & \mathbf{0} & 0 \\
\vdots & & \vdots & & \vdots & & \vdots & & & & \vdots \\
\mathbf{0} & 0 & \mathbf{0} & 0 & \dots & \frac{\partial \Delta \mathbf{V}_{n-1}}{\partial \mathbf{R}_{n-2}} & \frac{\partial \Delta \mathbf{V}_{n-1}}{\partial t_{n-2}} & \frac{\partial \Delta \mathbf{V}_{n-1}}{\partial \mathbf{R}_{n-1}} & \frac{\partial \Delta \mathbf{V}_{n-1}}{\partial t_{n-1}} & \frac{\partial \Delta \mathbf{V}_{n-1}}{\partial \mathbf{R}_n} & \frac{\partial \Delta \mathbf{V}_{n-1}}{\partial t_n}
\end{bmatrix}$$

$[M]$

(2.24)

In order to generate M , one requires knowledge of each of the partials given in Eq. (2.24). Fortunately, each of these partials may be constructed using the state transition matrix, provided that the linear approximations are acceptable. In order to derive the formulae needed to represent each of these partials, we must examine the problem definition more closely.

We again consider the first two segments, defined by the first three patchpoints: P_1 , P_2 , and P_3 . Each of these patchpoints is characterized by its position \mathbf{R} , velocity \mathbf{V} , and time t . After the application of the Level 1 correction, Segment 1 traverses from P_1 to P_2 and Segment 2 traverses from P_2 to P_3 . The resulting trajectory is continuous in position over time (within some small tolerance at P_2) and continuous in velocity over time except at P_2 , where $\Delta \mathbf{V}_2$ defines the difference between \mathbf{V}_2^+ (the velocity at the start of Segment 2) and \mathbf{V}_2^- (the velocity at the end of Segment 1)

$$\Delta \mathbf{V}_2 = \mathbf{V}_2^+ - \mathbf{V}_2^-$$

The superscripts “−” and “+” differentiate between the incoming and outgoing parameters, respectively, at a particular patchpoint. The position, velocity, and time of the end of Segment 1 are indicated as \mathbf{R}_2^- , \mathbf{V}_2^- , and t_2^- , respectively. Likewise, the position, velocity, and time of the initial state of Segment 2 are indicated as \mathbf{R}_2^+ , \mathbf{V}_2^+ , and t_2^+ , respectively. After applying Level 1 to P_1 , $\mathbf{R}_2^- = \mathbf{R}_2^+$ and $t_2^- = t_2^+$. These are fixed constraints and assumed in the formulation of Level 2.

The state transition matrix, Φ , may be mapped from P_1 to P_2 and from P_3 to P_2 to approximate the response of \mathbf{V}_2^- and \mathbf{V}_2^+ , respectively, given a change in the states of P_1 and P_3 . The basic Level 2 formulation defines the state transition matrix to be a 6×6 matrix as given in Eq. (2.18)

$$\begin{bmatrix} \delta \mathbf{R}_f \\ \delta \mathbf{V}_f \end{bmatrix} = \Phi \begin{bmatrix} \Delta \mathbf{R}_0 \\ \Delta \mathbf{V}_0 \end{bmatrix}$$

where we indicate a change in parameters performed by the user by a “ Δ ” and the response by a “ δ ”. The 6×6 state transition matrix may be broken up into four 3×3 submatrices as illustrated previously in Eq. (2.18) and repeated here

$$\begin{bmatrix} \delta \mathbf{R}_f \\ \delta \mathbf{V}_f \end{bmatrix} = \begin{bmatrix} \Phi_{RR}(t_f, t_0) & \Phi_{RV}(t_f, t_0) \\ \Phi_{VR}(t_f, t_0) & \Phi_{VV}(t_f, t_0) \end{bmatrix} \begin{bmatrix} \Delta \mathbf{R}_0 \\ \Delta \mathbf{V}_0 \end{bmatrix} \quad (2.25)$$

In order to simplify the nomenclature, we abbreviate the pieces of Eq. (2.25) as follows, where the state transition matrix is now demonstrating a mapping of the deviations from patchpoint P_1 to patchpoint P_2

$$\begin{array}{c} \delta \mathbf{R}_2^- \\ \delta \mathbf{V}_2^- \end{array} = \begin{array}{cc} A_{21} & B_{21} \\ C_{21} & D_{21} \end{array} \begin{array}{c} \Delta \mathbf{R}_1^+ \\ \Delta \mathbf{V}_1^+ \end{array} \quad (2.26)$$

This simplified notation is commonly found in literature [134]. Using this simplification, the matrix C_{23} would describe the change in \mathbf{V}_2^+ caused by a perturbation in the position of P_3 , namely, \mathbf{R}_3^- .

The Level 1 differential corrector process given in Eq. (2.19) may be written using this simplified notation as follows, where we have again applied it to describe the linear approximation of the change in velocity of P_1 needed to achieve a position difference at the time t_2

$$\Delta \mathbf{V}_1^+ = B_{21}^{-1} \delta \mathbf{R}_2^-$$

The following two linear systems represent approximations of the changes in P_2 that are caused by deviations in the patchpoints P_1 and P_3 [134], where it is assumed that the deviations are all small enough to be in the linear regime of the dynamics along each trajectory

$$\begin{array}{c} \delta \mathbf{R}_2^- - \mathbf{V}_2^- \delta t_2^- \\ \delta \mathbf{V}_2^- - \mathbf{a}_2^+ \delta t_2^- \end{array} = \begin{array}{cc} A_{21} & B_{21} \\ C_{21} & D_{21} \end{array} \begin{array}{c} \Delta \mathbf{R}_1^+ - \mathbf{V}_1^+ \Delta t_1^+ \\ \Delta \mathbf{V}_1^+ - \mathbf{a}_1^+ \Delta t_1^+ \end{array} \quad (2.27)$$

$$\begin{array}{c} \delta \mathbf{R}_2^+ - \mathbf{V}_2^+ \delta t_2^+ \\ \delta \mathbf{V}_2^+ - \mathbf{a}_2^+ \delta t_2^+ \end{array} = \begin{array}{cc} A_{23} & B_{23} \\ C_{23} & D_{23} \end{array} \begin{array}{c} \Delta \mathbf{R}_3^- - \mathbf{V}_3^- \Delta t_3^- \\ \Delta \mathbf{V}_3^- - \mathbf{a}_3^- \Delta t_3^- \end{array} \quad (2.28)$$

The formulation for this particular Level 2 differential corrector includes the fixed constraints that the trajectory be continuous in position and time across each patchpoint. Hence, $\mathbf{R}_2^- = \mathbf{R}_2^+ = \mathbf{R}_2$ and $t_2^- = t_2^+ = t_2$. For most applications, this also implies that $\mathbf{a}_2^- = \mathbf{a}_2^+ = \mathbf{a}_2$, though that may not be the case in the presence of dynamics that are velocity-dependent, such as atmospheric drag. These fixed constraints will be applied to each and every patchpoint in turn as the matrix M is constructed.

The targets for this Level 2 are that $\mathbf{V}_2^- = \mathbf{V}_2^+$ in order that the trajectory require no $\Delta \mathbf{V}$. The formulation is nearly identical for the case when a mission designer wishes to specify that a particular $\Delta \mathbf{V}$ be performed at a patchpoint. Hence, the more general target is given by

$$\Delta \hat{\mathbf{V}}_2 - \mathbf{V}_2^+ - \mathbf{V}_2^- = 0 \quad (2.29)$$

where the vector $\Delta \hat{\mathbf{V}}_2$ is specified by the designer.

As described earlier, the controls available to achieve the target $\Delta \hat{\mathbf{V}}_2$ include the position vectors and times of P_1 , P_2 , and P_3 . The controls and constraints applied to achieve the target $\Delta \hat{\mathbf{V}}_2$ permit \mathbf{V}_1^+ and \mathbf{V}_3^- to be free variables, though those may be targeted by neighboring constraints as the matrix M is constructed.

In summary, there are 28 parameters involved with the goal of achieving a desirable ΔV across P_2 , including the position, velocity, and time at the beginning and end of each trajectory segment, organized as follows

$$\begin{aligned} \text{Fixed Constraints: } & \mathbf{R}_2^- = \mathbf{R}_2^+ = \mathbf{R}_2, \quad t_2^- = t_2^+ = t_2 \\ \text{Controls: } & \Delta \mathbf{R}_1, \Delta t_1, \Delta \mathbf{R}_2, \Delta t_2, \Delta \mathbf{R}_3, \Delta t_3 \\ \text{Free Variables: } & \Delta \mathbf{V}_1^+, \Delta \mathbf{V}_3^- \\ \text{Targets: } & \delta \mathbf{V}_2^+ - \delta \mathbf{V}_2^- = \Delta \hat{\mathbf{V}}_2 - \mathbf{V}_2^+ - \mathbf{V}_2^- \end{aligned}$$

A similar set of parameters is defined for each patchpoint that is included in the differential correction process.

The first row of the M matrix requires six partial derivatives. These in turn require other partial derivatives, as follows

$$\begin{aligned} \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_1} &= -\frac{\partial \mathbf{V}_2^-}{\partial \mathbf{R}_1^+} & \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_2} &= \frac{\partial \mathbf{V}_2^+}{\partial \mathbf{R}_2^-} - \frac{\partial \mathbf{V}_2^-}{\partial \mathbf{R}_2^+} & \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_3} &= \frac{\partial \mathbf{V}_2^+}{\partial \mathbf{R}_3^-} \\ \frac{\partial \Delta \mathbf{V}_2}{\partial t_1} &= -\frac{\partial \mathbf{V}_2^-}{\partial t_1^+} & \frac{\partial \Delta \mathbf{V}_2}{\partial t_2} &= \frac{\partial \mathbf{V}_2^+}{\partial t_2^-} - \frac{\partial \mathbf{V}_2^-}{\partial t_2^+} & \frac{\partial \Delta \mathbf{V}_2}{\partial t_3} &= \frac{\partial \mathbf{V}_2^+}{\partial t_3^-} \end{aligned}$$

Wilson provides details to construct each of these partials [134]; we will demonstrate the process and illustrate the construction of $\frac{\partial \mathbf{V}_2^-}{\partial \mathbf{R}_1^+}$; the process may be applied in the same manner to construct each of these partials.

In order to construct $\frac{\partial \mathbf{V}_2^-}{\partial \mathbf{R}_1^+}$, we first set the perturbation of every other independent control to zero, namely

$$\begin{aligned} \delta \mathbf{R}_2 &= \delta \mathbf{R}_3 = \mathbf{0} \\ \delta t_1 &= \delta t_2 = \delta t_3 = 0 \end{aligned}$$

These values may then be inserted into Eqs. (2.27) and (2.28) or their inverses, whichever generates the most practical result. There are often many ways to describe the partials, and we are interested in the simplest relationships. For this particular case, the simplest relationship comes from substituting these values into the inverse of Eq. (2.27)

$$\begin{aligned} \delta \mathbf{R}_1^+ - \mathbf{V}_1^+ \delta t_1^+ \\ \delta \mathbf{V}_1^+ - \mathbf{a}_1^+ \delta t_1^+ \end{aligned} = \begin{bmatrix} A_{12} & B_{12} \\ C_{12} & D_{12} \end{bmatrix} \begin{aligned} \delta \mathbf{R}_2^- - \mathbf{V}_2^- \delta t_2^- \\ \delta \mathbf{V}_2^- - \mathbf{a}_2^- \delta t_2^- \end{aligned} \quad (2.30)$$

$$\begin{aligned} \delta \mathbf{R}_1^+ \\ \delta \mathbf{V}_1^+ \end{aligned} = \begin{bmatrix} A_{12} & B_{12} \\ C_{12} & D_{12} \end{bmatrix} \begin{aligned} \mathbf{0} \\ \delta \mathbf{V}_2^- \end{aligned} \quad (2.31)$$

This yields a system of two equations

$$\delta \mathbf{R}_1^+ = B_{12} \delta \mathbf{V}_2^- \quad \text{and} \quad (2.32)$$

$$\delta \mathbf{V}_1^+ = D_{12} \delta \mathbf{V}_2^- \quad (2.33)$$

The first equation provides the relationship we are interested in, namely

$$\frac{\delta \mathbf{V}_2^-}{\delta \mathbf{R}_1^+} = B_{12}^{-1} \quad (2.34)$$

Be aware that although $\Phi_{21} = \Phi_{12}^{-1}$, the submatrices do not typically follow such inverse relationships; that is, $B_{12}^{-1} \neq B_{21}$.

This procedure may be followed to generate relationships for each partial required for the production of the matrix M . The result is the following

$$\begin{aligned} \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_1} &= -\frac{\partial \mathbf{V}_2^-}{\partial \mathbf{R}_1^+} = -B_{12}^{-1} \\ \frac{\partial \Delta \mathbf{V}_2}{\partial t_1} &= -\frac{\partial \mathbf{V}_2^-}{\partial t_1^+} = B_{12}^{-1} \mathbf{V}_1^+ \\ \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_2} &= \frac{\partial \mathbf{V}_2^+}{\partial \mathbf{R}_2^-} - \frac{\partial \mathbf{V}_2^-}{\partial \mathbf{R}_2^+} = -B_{32}^{-1} A_{32} + B_{12}^{-1} A_{12} \\ \frac{\partial \Delta \mathbf{V}_2}{\partial t_2} &= \frac{\partial \mathbf{V}_2^+}{\partial t_2^-} - \frac{\partial \mathbf{V}_2^-}{\partial t_2^+} = [\mathbf{a}_2^+ - \mathbf{a}_2^-] + [B_{32}^{-1} A_{32} \mathbf{V}_2^+ - B_{12}^{-1} A_{12} \mathbf{V}_2^-] \\ \frac{\partial \Delta \mathbf{V}_2}{\partial \mathbf{R}_3} &= \frac{\partial \mathbf{V}_2^+}{\partial \mathbf{R}_3^-} = B_{32}^{-1} \\ \frac{\partial \Delta \mathbf{V}_2}{\partial t_3} &= \frac{\partial \mathbf{V}_2^+}{\partial t_3^-} = -B_{32}^{-1} \mathbf{V}_3^- \end{aligned}$$

Finally, we have all of the pieces to use Eq. (2.22) to determine an approximation of the adjustments that must be made in the positions and times of each patchpoint as a function of the unwanted velocity changes in each patchpoint.

Level 2 with Constraints. The Level 2 differential corrector can be modified to place a wide variety of constraints on the patchpoints in the system. For instance, we already observed in the derivation of the partials above that it is quite arbitrary to enforce the $\Delta \mathbf{V}$ at each patchpoint to zero; rather, one can specify a list of $\Delta \mathbf{V}$ values to perform at particular times and drive the trajectory to that solution instead. Before doing that, we must have a way of preventing the Level 2 corrector from adjusting a patchpoint's time. This is one example of a constraint that may be placed on the system. Other examples include constraining a patchpoint to have a particular inclination relative to some body, or to be located at a particular position or distance relative to a body. These constraints are very important when designing a practical trajectory for a spacecraft mission. For instance, the trajectory being designed may be an extension to a spacecraft's mission that is already in orbit, such that the trajectory must originate from the spacecraft's current trajectory. Or perhaps the trajectory being designed must land on the Moon at a particular landing site. The multiple-shooting differential corrector can accommodate any of these scenarios.

Any constraint may be added to the Level 2 architecture as long as it may be described in the form

$$\alpha_{ij} = f(\mathbf{R}_i, \mathbf{V}_i, t_i) \quad (2.35)$$

where the subscript i represents the patchpoint that the constraint is placed upon and the subscript j indicates the constraint number applied to that patchpoint. This nomenclature is consistent with that used by previous authors [134]. In this form, a constraint may be treated precisely the same as the ΔV targets described in the previous section. The constraint will be added to the list of targets for the differential corrector. It will be assumed, once again, that the only controls that may influence the constraint are the position and time of the patchpoint that the constraint is applied to as well as the positions and times of the two neighboring patchpoints (or the single neighboring patchpoint in the case of a constraint placed on the first or last patchpoint of a trajectory). The differential corrector may certainly be re-derived to operate with constraints that act upon many patchpoints, but this discussion is limited to constraints that act upon a single patchpoint.

It is straightforward to add a constraint to the list of targets in the differential corrector. The relationship given in Eq. (2.23) is augmented as follows

$$\begin{matrix} \delta \Delta \mathbf{V}_i \\ \delta \alpha_{ij} \end{matrix} = \underbrace{\begin{bmatrix} \frac{\partial \Delta \mathbf{V}_i}{\partial \mathbf{R}_k} & \frac{\partial \Delta \mathbf{V}_i}{\partial t_k} \\ \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_k} & \frac{\partial \alpha_{ij}}{\partial t_k} \end{bmatrix}}_{[P]} \begin{matrix} \delta \mathbf{R}_k \\ \delta t_k \end{matrix} \quad (2.36)$$

where the matrix P is known as the augmented state relationship matrix (SRM). Equation (2.36) is highly compressed: P is typically sparsely populated roughly along the diagonal, such that each constraint and each ΔV may only be influenced by the patchpoint it is assigned to and that patchpoint's nearest neighbors. Much like the ΔV constraints described in the previous section, each constraint requires the definition of the following six partials

$$\begin{matrix} \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_{i-1}} & \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_i} & \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_{i+1}} \\ \frac{\partial \alpha_{ij}}{\partial t_{i-1}} & \frac{\partial \alpha_{ij}}{\partial t_i} & \frac{\partial \alpha_{ij}}{\partial t_{i+1}} \end{matrix} \quad (2.37)$$

A quick observation shows that the ΔV constraints described in the previous section are a specific case of a constraint, where $\alpha_{ij} = \Delta \mathbf{V}_i$.

Any constraint that is a function of the position and/or time of one of the control patchpoints, and not a direct function of the velocity of any patchpoint, may be easily defined. For instance, if one wishes to constrain the time of patchpoint P_i , one simply characterizes that constraint as

$$\alpha_{ij} = t_i - \hat{t}_i$$

where \hat{t}_i is the desired time. One then computes the partials given in Eq. (2.37) and finds that the only non-zero partial is

$$\frac{\partial \alpha_{ij}}{\partial t_i} = 1$$

Similarly, if one wishes to constrain the position vector of patchpoint P_i , one characterizes that constraint as

$$\alpha_{ij} = \mathbf{R}_i - \hat{\mathbf{R}}_i$$

where $\hat{\mathbf{R}}_i$ is the desired position vector. One then finds that the only non-zero partial is

$$\frac{\partial \alpha_{ij}}{\partial \mathbf{R}_i} = I_{3 \times 3}$$

Constraints that depend on velocity are more complex, as demonstrated by the ΔV constraints given above. In order to compute the partials given in Eq. (2.37), one must perform the chain rule and compute additional partial derivatives. Fortunately, many of these were computed in the previous section, and many go to zero for numerous constraint formulations. The relationships are

$$\begin{aligned} \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_{i-1}} &= \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_{i-1}} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^-} \frac{\partial \mathbf{V}_i^-}{\partial \mathbf{R}_{i-1}} \\ \frac{\partial \alpha_{ij}}{\partial t_{i-1}} &= \frac{\partial \alpha_{ij}}{\partial t_{i-1}} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^-} \frac{\partial \mathbf{V}_i^-}{\partial t_{i-1}} \\ \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_i} &= \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_i} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^-} \frac{\partial \mathbf{V}_i^-}{\partial \mathbf{R}_i} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^+} \frac{\partial \mathbf{V}_i^+}{\partial \mathbf{R}_i} \\ \frac{\partial \alpha_{ij}}{\partial t_i} &= \frac{\partial \alpha_{ij}}{\partial t_i} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^-} \frac{\partial \mathbf{V}_i^-}{\partial t_i} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^+} \frac{\partial \mathbf{V}_i^+}{\partial t_i} \\ \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_{i+1}} &= \frac{\partial \alpha_{ij}}{\partial \mathbf{R}_{i+1}} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^+} \frac{\partial \mathbf{V}_i^+}{\partial \mathbf{R}_{i+1}} \\ \frac{\partial \alpha_{ij}}{\partial t_{i+1}} &= \frac{\partial \alpha_{ij}}{\partial t_{i+1}} + \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^+} \frac{\partial \mathbf{V}_i^+}{\partial t_{i+1}} \end{aligned}$$

Wilson derives the formulae that may be used to constrain a patchpoint's velocity, velocity magnitude, inclination, apse location, flight path angle, declination, right ascension, and conic energy [134]. For example, the conic energy relative to a massive body may be described as

$$\begin{aligned} \alpha_{ij} &= \frac{|\mathbf{V}_i|^2}{2} - \frac{\mu}{|\mathbf{R}_i|} \\ &= \frac{\mathbf{V}_i \cdot \mathbf{V}_i}{2} - \frac{\mu}{(\mathbf{R}_i \cdot \mathbf{R}_i)^{1/2}} \end{aligned}$$

where μ is the gravitational parameter for the central body. The majority of the partials given above are either zero or already known. The remaining partials may be computed as follows

$$\begin{aligned}\frac{\partial \alpha_{ij}}{\partial \mathbf{R}_i} &= \frac{\mu \mathbf{R}_i^T}{|\mathbf{R}_i|^3} \\ \frac{\partial \alpha_{ij}}{\partial \mathbf{V}_i^\pm} &= \mathbf{V}_i^{\pm T} \\ \frac{\partial \alpha_{ij}}{\partial t_i} &= 0\end{aligned}$$

The implementation of additional constraints is left to the designer.

A practical constraint that is not formulated in the same way is to restrict the size of the steps that the Level 2 differential corrector may take between iterations. The differential corrector estimates the change in each patchpoint's position, time, or both in order to achieve the given targets, and it does so using a large system of linearized equations. It is often the case that small perturbations drive the realized deviations in the trajectory into highly nonlinear regimes. In practice, it is often the case that the application of a full adjustment in the controls will push the trajectory further from the desired solution than it started. If a designer observes the trajectory diverging from the desired target, one common solution is to limit the maximum deviation that the patchpoints may shift in position or time per iteration of the differential corrector. If implemented properly, the smaller steps should keep the trajectory within the basin of convergence of the solution.

Level 2 with Multiple Trajectories. The Level 2 differential corrector formulated here operates on a large system of controls, targets, and constraints, where ultimately each patchpoint in the system contributes to the satisfaction of all goals, though each patchpoint is only directly influenced by its neighboring patchpoints at any given iteration. This system may be applied to multiple trajectories simultaneously in much the same way as it is applied to a single trajectory. This has clear practical applications for many spacecraft missions that involve deployments, separations, and/or formation-flying activities.

One may formulate the Level 2 differential corrector with multiple trajectories by augmenting the SRM, P , once again, such that it includes the patchpoints, targets, and constraints of every trajectory. One must be sure to permit the system some ΔV leverage to allow any given pair of trajectories to separate (either forward in time or backward in time, as appropriate).

For example, let us assume we have a scenario that involves one spacecraft deploying a secondary payload via a spring mechanism, which imparts a specified ΔV between the two spacecraft. Let us also assume that the differential corrector is permitted to vary the trajectory of the joined system prior to the deployment, as well as both trajectories after the deployment, and the deployment ΔV may occur in any direction. One way to model this scenario is to set up two series of patchpoints that define each spacecraft and then carefully lock the two spacecraft together. A practical

way to lock the two trajectories together is to define the first patchpoint of the deployed payload to be prior to the deployment, such that its position, velocity, and time are all constrained to be equal to the relevant parameters of the host spacecraft. That is, it is entirely constrained to match the corresponding patchpoint on the host spacecraft's trajectory. The second patchpoint in the deployed payload's trajectory is then defined to be the deployment event, such that its position and time are constrained to be equal to the position and time of the host spacecraft at the deployment, and its outgoing velocity is constrained to have the appropriate ΔV magnitude applied. From there the trajectory departs in the same way as any other trajectory. In this scenario, one would also have to take care to model the appropriate reaction to the host spacecraft's trajectory via constraints.

The augmented SRM for the case of multiple trajectories is very sparse, and it may be beneficial to implement numerical algorithms that take advantage of this feature. The simple Step 2 SRM includes nonzero elements only on a diagonal swath six elements wide. The SRM shifts further away from diagonal each time it is augmented by an additional constraint or an additional trajectory, though it remains approximately diagonal.

2.6.6 Constructing Periodic Orbits

Periodic orbits are important when analyzing and constructing trajectories using dynamical systems methods, since they help to characterize the flow in the system. There are many methods that are frequently used to identify and construct periodic orbits. Three categories of methods are described here:

1. **Analytical Expansion Techniques.** The discussion given in Section 2.6.2.2 demonstrates how to use basic analytical techniques to identify planar and three-dimensional periodic orbits in the CRTBP. Many authors have constructed analytical expansions that may be used to approximate periodic orbits in the CRTBP or in more complex systems [67, 123, 139].
2. **Shooting Techniques.** One may numerically construct a periodic orbit by targeting a single state as both the initial and final states in a trajectory using either a single- or multiple-shooting technique. This technique is difficult without any constraints, but it has proven to be very useful when numerically constructing certain types of periodic orbits, such as simple symmetric periodic orbits [46, 107, 108, 122].
3. **The Poincaré Method.** The Poincaré Method is a notable method that has proven to be very successful at identifying periodic and quasiperiodic orbits, especially stable orbits. Poincaré's technique involves numerically integrating many trajectories for a large amount of time. Trajectories that are close to periodic tend to linger near the same regions of the state space. One can readily identify stable periodic orbits or trajectories near such orbits if one places a plane in the state space, that is, a *Poincaré Surface of Section*. Then one records the state of each trajectory as the trajectory pierces the plane.

A periodic orbit appears as a fixed-point in the plane; a quasiperiodic orbit appears as a closed loop in the plane. Regions that are unstable in the state space appear as a chaotic sea of points, since unstable trajectories are very sensitive to their initial conditions.

Many other types of methods certainly exist, but these three categories provide a good overview of the variety of methods that are frequently used.

2.6.6.1 Periodic Orbits in the CRTBP If the Lagrange points represent the five simplest solutions to the CRTBP, it may be argued that the next set of solutions to introduce is the set of simple periodic symmetric orbits in the CRTBP. Simple periodic symmetric orbits are orbits that are symmetric about the $y = 0$ plane, pierce the $y = 0$ plane exactly twice per orbit, and pierce the plane orthogonally each time. Libration orbits, such as halo and Lyapunov orbits, are good examples of such orbits. A simple single-shooting differential correction scheme may be used to construct these orbits by taking advantage of their well-defined structure. Section 2.6.6.2 provides more information about this differential correction scheme. It should be noted that although this class of orbits does include what might be argued to be the simplest periodic orbits in the CRTBP, this class of orbits also includes families of very complex orbits.

Many other types of periodic orbits exist in the CRTBP, including orbits that pierce the $y = 0$ plane multiple times per orbit and orbits that are not symmetric, such as orbits about the triangular Lagrange points. One may also construct arbitrarily complex periodic orbits by chaining simple unstable orbits together, as is discussed in Section 2.6.11.

2.6.6.2 Single-Shooting Method for Constructing Simple Periodic Symmetric Orbits in the CRTBP One may formulate many types of shooting techniques to identify periodic orbits using the techniques introduced in Section 2.6.5. Howell identified a simple procedure that has been used by many researchers in the field [122]. The technique is easily applied to the families of halo orbits, Lyapunov orbits, distant prograde orbits, distant retrograde orbits, symmetric resonant orbits, and a variety of other classes of symmetric periodic orbits [46]. Since it is a very common and straightforward procedure, and since it has been used repeatedly in relevant research to construct halo orbits and other similar orbits, it is reviewed here.

As mentioned earlier, simple periodic symmetric orbits are orbits that are symmetric about the $y = 0$ plane, pierce the $y = 0$ plane exactly twice per orbit, and pierce the plane orthogonally each time. Let us define $\mathbf{X}(t_0)$ to be the state of a simple periodic symmetric orbit at the $y = 0$ plane-crossing with a positive \dot{y} and $\mathbf{X}(t_{T/2})$ to be the state of the orbit half of its orbital period later at the $y = 0$ plane-crossing with a negative \dot{y} . For this orbit to be periodic and symmetric, these states must have the following form

$$\begin{aligned}\mathbf{X}(t_0) &= [x_0 \ 0 \ z_0 \ 0 \ \dot{y}_0 \ 0]^T \\ \mathbf{X}(t_{T/2}) &= [x_{T/2} \ 0 \ z_{T/2} \ 0 \ \dot{y}_{T/2} \ 0]^T\end{aligned}\tag{2.38}$$

Let us assume that we have an initial guess, $\hat{\mathbf{X}}(t_0)$, that is near the initial state of a desirable orbit. When we integrate this state forward in time until the next $y = 0$ plane, we obtain the state $\hat{\mathbf{X}}(t_{\hat{T}/2})$

$$\hat{\mathbf{X}}(t_{\hat{T}/2}) = \begin{bmatrix} x_{\hat{T}/2} & 0 & z_{\hat{T}/2} & \dot{x}_{\hat{T}/2} & \dot{y}_{\hat{T}/2} & \dot{z}_{\hat{T}/2} \end{bmatrix}^T$$

We now wish to adjust the initial state of the trajectory in such a way as to drive the values of $\dot{x}_{\hat{T}/2}$ and $\dot{z}_{\hat{T}/2}$ to zero. One notices that by adjusting the initial state, not only do the values of \dot{x} and \dot{z} change, but the propagation time, $\hat{T}/2$, needed to pierce the $y = 0$ plane also changes. In order to target a proper state $\mathbf{X}(t_{T/2})$, one may vary the initial values of x , z , and/or \dot{y} .

The linearized system of equations relating the final state to the initial state may be written as

$$\delta\mathbf{X}(t_{T/2}) \approx \Phi(t_{T/2}, t_0) \delta\mathbf{X}(t_0) + \frac{\partial\mathbf{X}}{\partial t} \delta(T/2) \quad (2.39)$$

where $\delta\mathbf{X}(t_{T/2})$ is the deviation in the final state due to a deviation in the initial state, $\delta\mathbf{X}(t_0)$, and a corresponding deviation in the orbit's period, $\delta(T/2)$. The time-derivative of the state, $\partial\mathbf{X}/\partial t$, may be computed at the second plane-crossing, namely, at time $t = T/2$. Equation (2.39) may be used as the driver for a differential corrector by setting $\delta\mathbf{X}(t_{T/2})$ to be the desired change in the final state's components and solving for $\delta\mathbf{X}(t_0)$, the approximate correction to the initial state needed to produce such a change.

We now consider what the desired change in the final state's components should be. For our purposes, the only desired change in the final state is a change in the values of \dot{x} and \dot{z} , but it is not important if the other components of the final state change. However, we know that the deviation in the final value of y will always be equal to zero since the trajectory is always propagated to that point. Thus we set $\delta\mathbf{X}(t_{T/2})$ to

$$\delta\mathbf{X}(t_{T/2}) = \begin{bmatrix} \delta x_{T/2} & 0 & \delta z_{T/2} & -\dot{x}_{T/2} & \delta\dot{y}_{T/2} & -\dot{z}_{T/2} \end{bmatrix}^T$$

Furthermore, in order to restrict our search to simple periodic symmetrical orbits, we restrict the allowed correction in the initial conditions to

$$\delta\mathbf{X}(t_0) = \begin{bmatrix} \delta x_0 & 0 & \delta z_0 & 0 & \delta\dot{y}_0 & 0 \end{bmatrix}^T$$

Now Eq. (2.39) simplifies to

$$\begin{bmatrix} \delta x_{T/2} \\ 0 \\ \delta z_{T/2} \\ -\dot{x}_{T/2} \\ \delta\dot{y}_{T/2} \\ -\dot{z}_{T/2} \end{bmatrix} \approx \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} & \phi_{16} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} & \phi_{26} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} & \phi_{36} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} & \phi_{46} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} & \phi_{56} \\ \phi_{61} & \phi_{62} & \phi_{63} & \phi_{64} & \phi_{65} & \phi_{66} \end{bmatrix} t_{T/2}, t_0 \begin{bmatrix} \delta x_0 \\ 0 \\ \delta z_0 \\ 0 \\ \delta\dot{y}_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \dot{z} \\ \dot{y} \\ \dot{x} \end{bmatrix} \delta(T/2) \quad (2.40)$$

The value of $\delta(T/2)$ may be determined from the second line of Eq. (2.40) to be

$$\delta(T/2) = \frac{-\phi_{21}\delta x_0 - \phi_{23}\delta z_0 - \phi_{25}\delta\dot{y}_0}{\dot{y}} \quad (2.41)$$

Substituting this value into the fourth and sixth lines of Eq. (2.40) yields

$$-\dot{x}_{T/2} \approx \phi_{41} - \phi_{21} \frac{\ddot{x}}{\dot{y}} \delta x_0 + \phi_{43} - \phi_{23} \frac{\ddot{x}}{\dot{y}} \delta z_0 + \phi_{45} - \phi_{25} \frac{\ddot{x}}{\dot{y}} \delta \dot{y}_0 \quad (2.42)$$

$$-\dot{z}_{T/2} \approx \phi_{61} - \phi_{21} \frac{\ddot{z}}{\dot{y}} \delta x_0 + \phi_{63} - \phi_{23} \frac{\ddot{z}}{\dot{y}} \delta z_0 + \phi_{65} - \phi_{25} \frac{\ddot{z}}{\dot{y}} \delta \dot{y}_0 \quad (2.43)$$

Equations (2.42) and (2.43) give expressions for the approximate deviation in the final x - and z -velocities as functions of the deviation in all three initial conditions x_0 , z_0 , and \dot{y}_0 . It is sufficient to change only two of the initial conditions, if that is desirable, or a combination of all three. For the purpose of this description, the value of x_0 will be kept constant, and the values of z_0 and \dot{y}_0 will be permitted to vary. The following expression summarizes the approximate changes that must be made to z_0 and \dot{y}_0 to produce a desirable change in the final state (while keeping the other initial conditions constant)

$$\begin{bmatrix} \delta z_0 \\ \delta \dot{y}_0 \end{bmatrix} \approx \begin{bmatrix} \phi_{43} - \phi_{23} \frac{\ddot{x}}{\dot{y}} & \phi_{45} - \phi_{25} \frac{\ddot{x}}{\dot{y}} \\ \phi_{63} - \phi_{23} \frac{\ddot{z}}{\dot{y}} & \phi_{65} - \phi_{25} \frac{\ddot{z}}{\dot{y}} \end{bmatrix}^{-1} \begin{bmatrix} -\dot{x}_{T/2} \\ -\dot{z}_{T/2} \end{bmatrix} \quad (2.44)$$

Since the system was linearized in order to produce this procedure, the adjustments will not correct the unwanted motion perfectly; this procedure must be iterated until it converges on an orbit.

When all is said and done, a simple, symmetric periodic orbit has three nonzero states at its orthogonal $y = 0$ plane crossing: x_0 , z_0 , and \dot{y} (see Eq. (2.38)). The procedure outlined here is used to generate the periodic orbit given one of those parameters and estimates of the other two. Because of this, a family of periodic orbits may be well represented by plotting its initial \dot{y} values or its Jacobi constant values as a function of its initial x values. Figure 2-14 illustrates these curves using the family of Lyapunov orbits about LL_1 as an example.

2.6.6.3 Differentially Correcting Orbits into the DE421 Model An orbit that is perfectly periodic in the CRTBP is not perfectly periodic in the real Solar System since the planets and moons in the real Solar System do not move in circular, coplanar orbits. Various perturbations lead the orbit to diverge from being periodic; the most notable of which is the nonzero eccentricity of the orbits of the primary bodies in the system [100].

To produce a quasiperiodic orbit in the real Solar System, one can use a multiple-shooting differential corrector with the periodic CRTBP orbit as the initial guess of the real trajectory. This technique was demonstrated in Section 2.6.5.2. The differential corrector takes the CRTBP orbit and perturbs it to keep it near its initial guess while eliminating the need to perform large maneuvers. In the case of generating a quasiperiodic halo or Lissajous orbit in the DE421 model of the Solar System, one may use an analytical approximation of the orbit as the initial guess to the differential corrector [123, 139]. This has been demonstrated on many occasions and has been shown to work well [47].

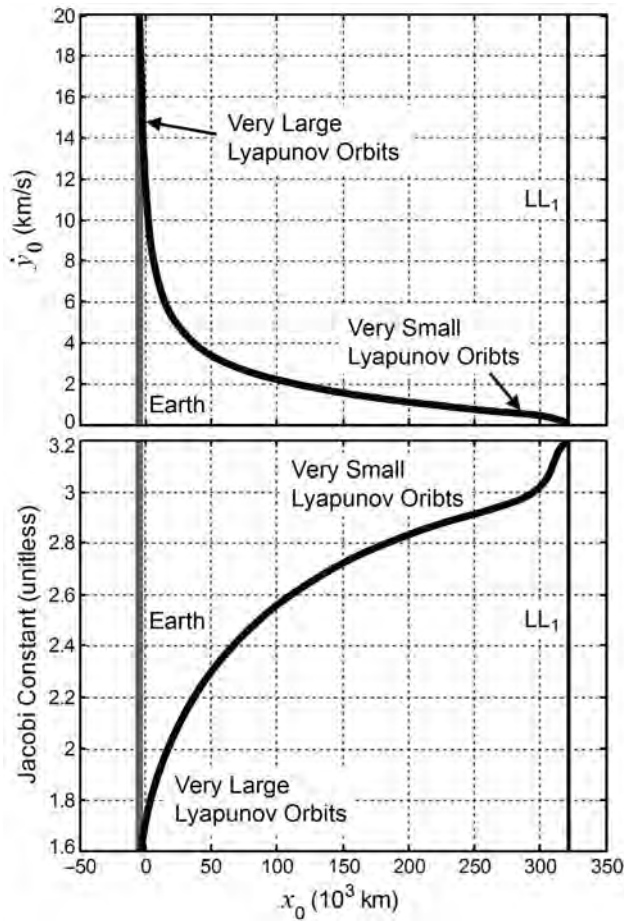


Figure 2-14 Plots of x_0 vs. \dot{y}_0 (top) and x_0 vs. C (bottom) for the family of Lyapunov orbits about the Earth–Moon L_1 point. The initial values of the other Cartesian coordinates in the Earth–Moon synodic frame are all equal to zero for each orbit in this family [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

Figure 2-15 shows the difference between a halo orbit about the lunar L_2 point produced in the CRTBP compared with the same halo orbit differentially corrected into the DE421 model of the real Solar System. One can see that the real halo orbit is quasiperiodic, tracing out the same vicinity of space on each orbit, but never truly retracing itself. For this illustration, the realistic quasi-halo orbit is plotted in a coordinate frame that is normalized over time based on the instantaneous distance between the Earth and the Moon, and then re-scaled to the average distance between the Earth and the Moon.

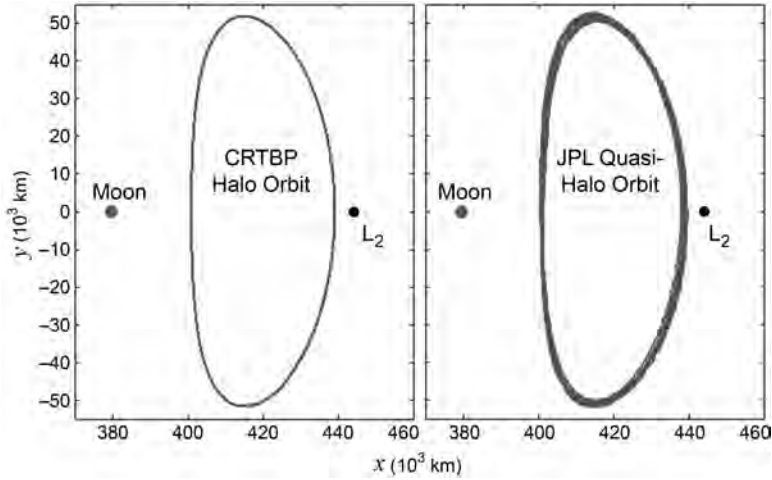


Figure 2-15 A comparison between a halo orbit produced in the CRTBP and a quasi-halo orbit produced in the DE421 model of the real Solar System. The orbits are shown in the Earth–Moon synodic reference frame [44, 46].

The perfectly periodic CRTBP orbit is typically a very close approximation of the real quasiperiodic orbit, enough so that the early mission design may be developed in the CRTBP. This is convenient because the motion of a spacecraft in a perfectly periodic orbit is more predictable than the motion of a spacecraft in a quasiperiodic orbit.

On several occasions, it has been observed that some of the structure of a periodic orbit in the CRTBP becomes lost or significantly altered as the orbit is differentially corrected into the DE421 model. This is often seen when a single revolution of a periodic orbit is sent into the differential corrector. Ordinarily, a differential corrector converges on a continuous trajectory more readily if the trajectory’s endpoints are not constrained. Without the boundary values constrained, it is often the case that the differential corrector significantly alters the states of the trajectory’s endpoints. The resulting trajectory, although continuous, may not resemble the original orbit much at all. This effect may be observed in Fig. 2-16.

One way to combat this effect is to differentially correct several orbits of the periodic orbit together. For the purpose of this discussion, let us say that four periodic orbits are differentially corrected together. Then, two of the orbits are “outer” orbits (the first and last orbits) that are vulnerable to substantial changes in the differential correction process, and two of the orbits are “inner” orbits (the second and third orbits) that are more protected from significant alteration in the process. Normally, the differential corrector converges on a continuous trajectory before the inner orbits are substantially altered. Once the differential corrector has converged on the final trajectory, then the outer orbits may be pruned off in order to observe the structure of

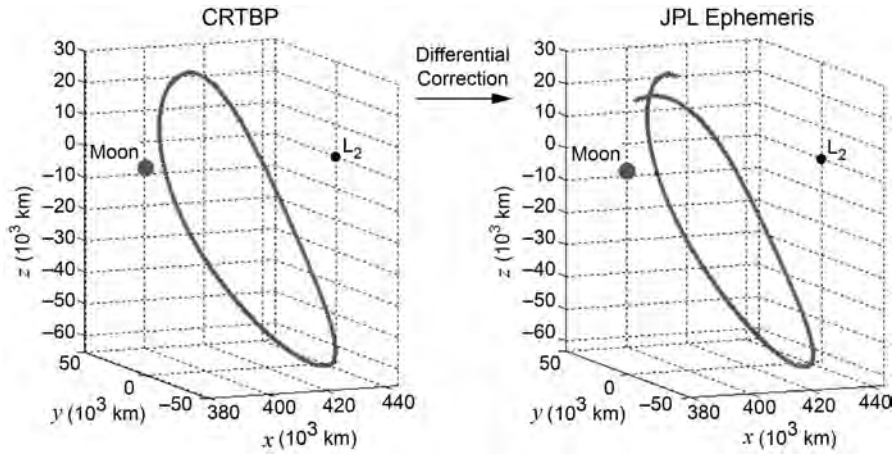


Figure 2-16 A single L_2 halo orbit in the Earth–Moon CRTBP (left) is differentially corrected into the DE421 model (right) [46].

the resulting quasiperiodic orbit. Ordinarily, this procedure results in quasiperiodic orbits that exist in the DE421 model that retain the same structure as the periodic orbits that exist in the CRTBP. Figure 2-17 shows an example of this process.

Since halo orbits are used frequently in later chapters of this book, some discussion is given here regarding the largest observable deviations between the perfectly periodic halo orbit in the CRTBP and the quasi-halo orbit in the real Solar System. Arguably the most substantial deviation between the CRTBP and the real Solar Sys-

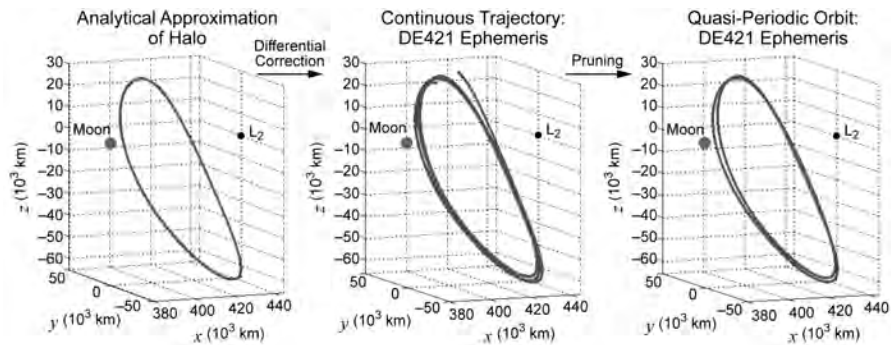


Figure 2-17 An example of the process of differentially correcting and pruning a halo orbit from the Earth–Moon CRTBP into the DE421 model. Left: the nominal periodic halo orbit in the Earth–Moon CRTBP; center: the differentially corrected trajectory in the DE421 model; right: the pruned quasiperiodic halo trajectory in the DE421 model [44, 46].

tem, at least in the case of the Sun–Earth and Earth–Moon three-body systems, is the nonzero eccentricity of the orbits of the primary bodies in the system. The real, eccentric orbit of the primaries imparts a deviation in the quasi-halo orbits that has a period equal to the orbital period of the primaries. Since most halo orbits have a period equal to approximately half of the orbital period of the primaries, this dynamic perturbation tends to appear as a resonant pulsation. One quasi-halo revolution tends to deviate from the perfectly periodic halo orbit in one direction, and the next revolution tends to deviate in the opposite direction. The result is that a spacecraft on a quasi-halo orbit tends to retrace its path very closely every other revolution. This effect is less visible when the reference frame is centered on a Lagrange point rather than the barycenter, since the Lagrange point pulses in and out as the primary bodies follow their noncircular orbits.

Figure 2-18 illustrates the pulsation that exists in the Earth–Moon system by showing a plot of the distance between the Moon and a spacecraft traversing an orbit much like that one shown in Figs. 2-16 and 2-17. One can see that every other revolution retraces a similar path. The moments in time when the Moon reaches its perigee and apogee are indicated for reference. Figure 2-19 illustrates how this same quasi-halo orbit appears in the DE421 ephemeris when viewed in different synodic coordinate systems, including an Earth-centered synodic frame, out to an LL_2 -centered synodic frame.

2.6.7 The Continuation Method

Periodic orbits in the CRTBP may be grouped into families, where a family consists of an infinite number of periodic orbits whose properties vary continuously from one end of the family to the other. All orbits in the same family may be uniquely identified by a single parameter of that family, for example, their position on a perpendicular $y = 0$ plane crossing, their velocity at that crossing, or some other specified parameter. This property of the CRTBP is due to the existence of the Jacobi

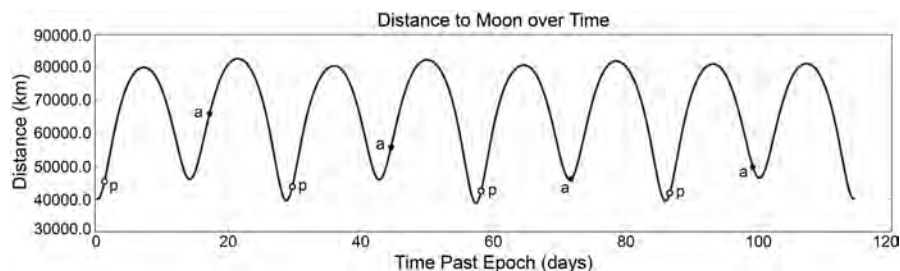


Figure 2-18 The distance between the orbit and the Moon over time for a realistic quasi-halo orbit. The moments in time when the Moon reaches its perigee and apogee are indicated by the symbols “p” and “a”, respectively.

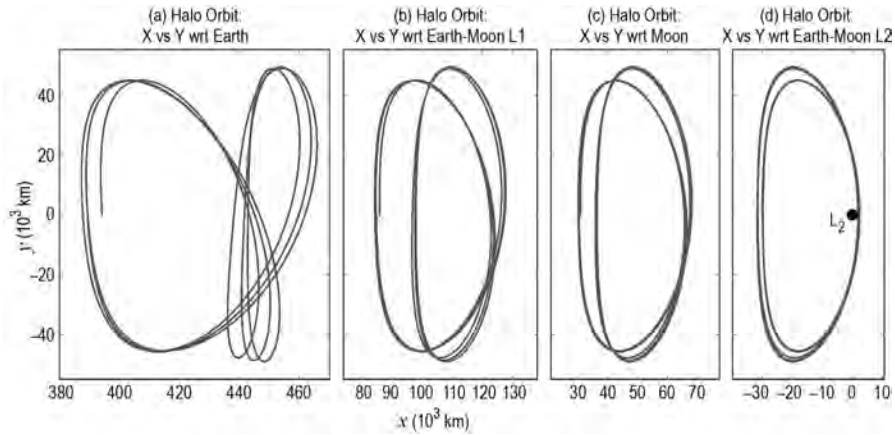


Figure 2-19 An illustration of how the same quasi-halo orbit appears in different synodic coordinate systems. From left to right, the systems include an Earth-centered synodic frame, an L_1 -centered frame, a Moon-centered frame, and an L_2 -centered frame.

constant, the CRTBP's unique integral of motion. Hénon provides more discussion about the existence of families of solutions in the CRTBP [113].

Once a single periodic orbit is known in the CRTBP, then the continuation method may be used to traverse that orbit's family. The method starts by perturbing some parameter of the known periodic orbit and then differentially correcting the new conditions to find that periodic orbit's neighbor in its family. The differential corrector presented in Section 2.6.6 is well-suited to this method for simple periodic symmetrical orbits because one may vary the initial position and correct for the initial velocity that corresponds to the next periodic orbit in the family (or vice versa, if desired).

To demonstrate this method, the continuation method has been applied to the family of Lyapunov orbits that exist about the Earth–Moon L_2 point. First, a single Lyapunov orbit is identified, for example, the gray orbit in Figs. 2-20 and 2-21. The orbit's initial position, x_0 , is then systematically varied while a differential corrector fills out the curve shown in Fig. 2-20. The initial conditions in the curve correspond to the family of orbits shown in Fig. 2-21.

The continuation method works well when the perturbations are small; in practice it is beneficial to predict the differential corrector's adjustment to the perturbation because this allows larger jumps in the varying parameter. Furthermore, if the perturbations are too large, the differential corrector may converge on a solution of a different family. Thus smaller steps or better prediction methods may be required to make the continuation method more reliable. The work for this study has implemented a quadratic prediction method that uses the three previous data points of the family to predict the next data point. This has been sufficient to allow the differential corrector to converge quickly while allowing the curve of the family to evolve naturally over the state space. Two-dimensional curve tracking algorithms may also work well since

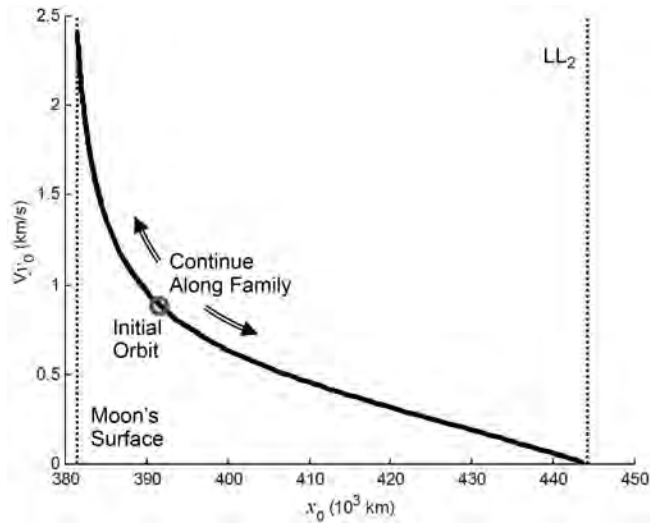


Figure 2-20 A plot of the initial conditions of the family of Lyapunov orbits about the Earth–Moon L_2 point (LL_2) [140] (Copyright ©2006 by American Astronautical Society Publications Office, San Diego, California [website <http://www.univelt.com>], all rights reserved; reprinted with permission of the AAS).

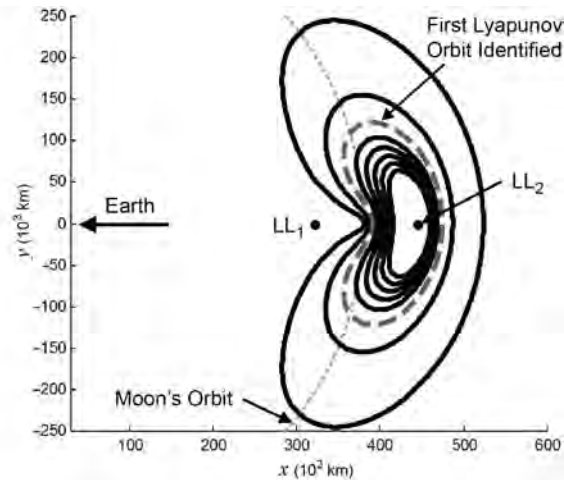


Figure 2-21 Plots of the orbits in the family of LL_2 Lyapunov orbits corresponding to those initial conditions shown in Fig. 2-20 [140] (Copyright ©2006 by American Astronautical Society Publications Office, San Diego, California [website <http://www.univelt.com>], all rights reserved; reprinted with permission of the AAS).

state space curves are not necessarily well-modeled by polynomials. For instance, one may extrapolate curves using a constant arc-length of two parameters [141].

2.6.8 Orbit Stability

The stability of a periodic orbit may be determined by analyzing the eigenvalues of the orbit's monodromy matrix. A random perturbation in the state of a spacecraft on an unstable orbit will cause the spacecraft's state to exponentially diverge from that of the original orbit over time; hence, the monodromy matrix of an unstable orbit includes at least one eigenvalue for which the real component is outside of the unit circle. This section explores the stability characteristics of periodic orbits via the eigenvalues of their monodromy matrices.

2.6.8.1 Eigenvalues of an Orbit's Monodromy Matrix The monodromy matrices of orbits in the CRTBP have six eigenvalues, λ_i for $i = 1, 2, \dots, 6$, corresponding to the eigenvectors \mathbf{v}_i . The eigenvalues of the monodromy matrix occur in reciprocal pairs [142], which is a direct consequence of the symplectic nature of the monodromy matrix, and of the state transition matrix in the CRTBP in general [143]. Additionally, a pair of eigenvalues of the monodromy matrix will be equal to unity because of the Jacobi integral of motion in the CRTBP [131, 142]. The eigenvalues are thus related in the following way

$$\lambda_2 = \frac{1}{\lambda_1} \quad \lambda_4 = \frac{1}{\lambda_3} \quad \lambda_5 = \lambda_6 = 1 \quad (2.45)$$

The monodromy matrices of periodic orbits in the planar CRTBP only have four eigenvalues: $(\lambda_1, 1/\lambda_1, 1, 1)$. Since those orbits may be computed in the spatial CRTBP by setting their z - and \dot{z} -components to zero, the remainder of this section only considers orbits in the full three-dimensional system.

The eigenvalues of the monodromy matrix of a periodic orbit in the CRTBP are the roots of a characteristic equation; furthermore, each has a characteristic exponent, α , where $\lambda = e^{\alpha T}$ and T is the period of the orbit. Then, the reciprocal of that eigenvalue is equal to: $1/\lambda = e^{-\alpha T}$. The characteristic exponents are sometimes referred to as Lyapunov characteristic exponents [144].

The monodromy matrices of Keplerian orbits, such as low Earth orbits (LEOs), have three pairs of eigenvalues that are all equal to 1, indicating that after a full orbit any given perturbation neither grows nor decays exponentially. The monodromy matrices of periodic orbits in the CRTBP may have other eigenvalue pairs, including real values not equal to 1 and pairs of complex numbers. Table 2-5 provides a summary of the resulting motion of a spacecraft in a periodic orbit, whose state is perturbed along the eigenvector corresponding to any type of given eigenvalue.

The stability of a periodic orbit may be identified by observing the resulting motion of a perturbed particle in that orbit or by computing the eigenvalues of the orbit's monodromy matrix and comparing those eigenvalues to the results given in Table 2-5. The following classification scheme for an orbit's stability is used in this work:

Table 2-5 A summary of the resulting motion of a spacecraft in a periodic orbit, whose state is perturbed along the eigenvector corresponding to a given eigenvalue. The result of a perturbation along the eigenvector corresponding to a complex eigenvalue includes a combination of the imaginary result and one of the real results listed.

Eigenvalue	Result of the perturbation
Real, within the range $[-1,1]$	The perturbation exponentially decays.
Real, equal to 1 or -1	The perturbation neither exponentially decays nor grows.
Real, outside of the range $[-1,1]$	The perturbation exponentially grows.
<hr/>	
Imaginary	After each orbital period, the perturbation oscillates about the spacecraft's original state.

- If an eigenvalue exists whose real component is outside of the range $[-1,1]$, then the periodic orbit is asymptotically unstable, referred to here as *unstable*, along the corresponding eigenvector.
- If the real component of each and every eigenvalue of the monodromy matrix is between -1 and 1 , then the orbit is *stable*. Given the relationships in Eq. (2.45), an orbit may only be stable in the CRTBP if each and every eigenvalue pair is complex with real components in the range $[-1,1]$.
- If the orbit is not unstable and there is at least one eigenvalue pair whose real component is equal to 1 , then the periodic orbit is *neutrally stable*, or a *center* [130].

Since every periodic orbit in the CRTBP has at least one pair of eigenvalues with values are equal to unity, then it is customary to ignore that pair of eigenvalues when classifying the stability of the orbit [108, 145].

To determine the eigenvalues of the monodromy matrix, it is useful to consider the characteristic equation, since many of the roots of this equation are already known

$$\begin{aligned}\det(M - \lambda I) &= (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)(\lambda - \lambda_4)(\lambda - \lambda_5)(\lambda - \lambda_6) = 0 \\ &= (\lambda - 1)^2(\lambda - \lambda_1)(\lambda - 1/\lambda_1)(\lambda - \lambda_3)(\lambda - 1/\lambda_3) = 0\end{aligned}\quad (2.46)$$

The relationship given in Eq. (2.46) may be re-written in terms of the new parameters p and q , keeping consistent with the nomenclature found in the literature [116]

$$(\lambda - 1)^2 \lambda^2 + p\lambda + 1 \quad \lambda^2 + q\lambda + 1 = 0 \quad (2.47)$$

Thus, $p = -(\lambda_1 + 1/\lambda_1)$ and $q = -(\lambda_3 + 1/\lambda_3)$. Equation (2.47) may also be factored in the following manner

$$(\lambda - 1)^2 \lambda^4 + (p + q)\lambda^3 + (pq + 2)\lambda^2 + (p + q)\lambda + 1 = 0 \quad (2.48)$$

Equation (2.48) may be re-written using the new parameters α , β , and γ , once again to keep consistent with the nomenclature found in the literature [116] (where α should not be confused with the characteristic exponent that corresponds to each eigenvalue)

$$(\lambda - 1)^2 \lambda^4 + \alpha \lambda^3 + \beta \lambda^2 + \alpha \lambda + \gamma = 0 \quad (2.49)$$

In this form it is clear that $\alpha = p + q$, $\beta = pq + 2$, and $\gamma = 1$. The benefits of factoring the characteristic equation into the parameters α , β , and γ arises at this point. Bray and Goudas derive a fast and simple method to compute α and β using the monodromy matrix [116, 145]

$$\alpha = 2 - \text{trace}(M) \quad (2.50)$$

$$\beta = \frac{\alpha^2 - \text{trace}(M^2)}{2} + 1 \quad (2.51)$$

It is then simple to determine the parameters p and q using knowledge of α and β

$$\left. \begin{matrix} p \\ q \end{matrix} \right\} = \frac{\alpha \pm \sqrt{\alpha^2 - 4\beta + 8}}{2} \quad (2.52)$$

It then follows that with knowledge of p and q one may determine the corresponding eigenvalues

$$\left. \begin{matrix} \lambda_1 \\ 1/\lambda_1 \end{matrix} \right\} = \frac{-p \pm \sqrt{p^2 - 4}}{2} \quad (2.53)$$

$$\left. \begin{matrix} \lambda_3 \\ 1/\lambda_3 \end{matrix} \right\} = \frac{-q \pm \sqrt{q^2 - 4}}{2} \quad (2.54)$$

The final two eigenvalues have already been predetermined and are given in Eq. (2.45) as $\lambda_5 = \lambda_6 = 1$. Thus, Eqs. (2.50)–(2.54) provide a fast and simple method to compute the six eigenvalues of the monodromy matrix. The corresponding eigenvectors may be computed in any standard way using the equation $M\mathbf{v}_i = \lambda_i\mathbf{v}_i$. It should be noted that the stable and unstable eigenvalues, λ^S and λ^U , of an orbit's monodromy matrix, are equal to the pair of real eigenvalues with the smallest and largest values, respectively, if they exist.

2.6.8.2 The Stability Index An orbit's *stability index* is defined in various ways in the literature depending on the author. Several authors, for example, Broucke [108], define the stability of a periodic orbit on the value of k , where k is equal to the sum of the real eigenvalues of the orbit. If $|k| > n$, where n is equal to the number of real eigenvalues in the orbit's monodromy matrix, then the orbit is unstable; if $|k| < n$, the orbit is stable; otherwise $|k| = n$ and the orbit is neutrally stable. One problem with such a definition is that the value of n may change depending on the orbit.

Another definition of the stability index is defined by Howell, among others, as follows [122]. If one considers the definition $k_i = \lambda_i + 1/\lambda_i$, one notices several things. First, the values of k_i may be easily computed using the parameters p and

q that were introduced above, namely: $k_1 = -p$, $k_2 = -q$, and $k_3 = 2$. Next, the value of k_i is always real and in the range $-2 \leq k \leq 2$ for stable orbits since the sum of a real pair is real and the sum of a complex conjugate pair is also real. Furthermore, if $|k_i| > 2$, then the real component of at least one of the eigenvalues summed is greater than 1 and the orbit is unstable. Since two of the eigenvalues of the orbit's monodromy matrix are equal to unity and it is conventional to ignore them, the stability index, k , may then be given by

$$k = \sup\{|\operatorname{Re}(k_1)|, |\operatorname{Re}(k_2)|\} \quad (2.55)$$

where the operator $\operatorname{Re}()$ only observes the real component of the operand. We have the final test: if $k > 2$ the orbit is unstable, if $k = 2$ the orbit is neutrally stable, and if $k < 2$ the orbit is stable.

2.6.8.3 The Perturbation Doubling Time The stability index defined in Eq. (2.55) certainly provides information about the stability of the orbit in question. However, it only provides limited information about the relative stability of different orbits. A highly unstable orbit may appear to be more stable than a weakly unstable orbit if the weakly unstable orbit's period is much greater than the highly unstable orbit's period. It is now of interest to find a parameter that may be used to directly compare the stability of two orbits regardless of their relative orbital periods.

The eigenvalues of the monodromy matrix of a periodic orbit are a function of the orbit's period, T , and a characteristic exponent, α , as follows

$$\lambda = e^{\alpha T} \quad (2.56)$$

To compare the stability of several orbits directly, one may either normalize the eigenvalues of the monodromy matrices or, equivalently, compare the characteristic exponents in some way.

An intuitive measure for comparison is the orbit's *perturbation doubling time* (for unstable orbits) or the orbit's *perturbation half-life* (for stable orbits). In this work, we refer to this time measurement as $\hat{\tau}$ for two reasons: first, to indicate that it is a normalized measurement and second, to distinguish it from the parameter τ that is used to identify points along an orbit (see Section 2.6.2.3). Given a spacecraft in an unstable orbit, the perturbation doubling time characterizes the length of time that is required for a perturbation in the spacecraft's state to double in magnitude. Similarly, given a spacecraft in a stable orbit, the perturbation half-life characterizes the length of time that is required for a perturbation in the spacecraft's state to be reduced by one half. For simplicity, we refer to this time measurement only as the perturbation doubling time, since it is generally more useful when designing real missions to compare this time measurement for unstable orbits.

After determining the eigenvalues of the orbit's monodromy matrix, one may use Eq. (2.56) to determine the corresponding characteristic exponents. If a spacecraft's state is perturbed at time $t = t_0$ from its nominal state by a perturbation with magnitude $\delta(t_0)$ along the eigendirection corresponding to the characteristic exponent α , the perturbation magnitude grows over time by the following expression

$$\delta(t) = \delta(t_0)e^{\alpha(t-t_0)} \quad (2.57)$$

Given a random perturbation in the spacecraft's state, the spacecraft's deviation over time is dominated by the component of that deviation that exists in the most unstable eigendirection, namely, by the direction indicated by the unstable eigenvalue λ^U . The perturbation doubling time may be computed by identifying the time, $t = t_0 + \hat{\tau}$, when the spacecraft's perturbed state is twice as far from its nominal position compared to its perturbed state at time $t = t_0$. One can find the perturbation doubling time by solving for $\hat{\tau}$ in Eq. (2.58), derived as follows

$$\begin{aligned}\delta(t) &= \delta(t_0)e^{\alpha(t-t_0)} \\ 2\delta(t_0) &= \delta(t_0)e^{\alpha\hat{\tau}} \\ 2 &= e^{\alpha\hat{\tau}}\end{aligned}\tag{2.58}$$

where α is the characteristic exponent that corresponds to the unstable eigenvalue, λ^U , of the orbit's monodromy matrix. The value of α may be computed using the simple relationship $\alpha T = \ln \lambda^U$, derived from Eq. (2.56). Hence, the time duration $\hat{\tau}$ may be computed using the expression

$$\hat{\tau} = \frac{\ln 2}{\ln \lambda^U} T\tag{2.59}$$

2.6.9 Examples of Practical Three-Body Orbits

The three-body problem contains a wide variety of interesting and potentially useful periodic and quasiperiodic orbits. Numerous authors have catalogued families of orbits and a brief history of these efforts is given in Section 2.6.2.2. This section illustrates several example families of three-body orbits, all of which appear frequently in the literature, and often in spacecraft mission proposals.

2.6.9.1 Lyapunov Orbits Lyapunov orbits were introduced in Section 2.6.2.2; they are two-dimensional periodic solutions to the circular restricted three-body problem. Lyapunov orbits exist about all three of the collinear Lagrange points, as illustrated in Fig. 2-22. The LL_1 and LL_2 families include orbits with orbital periods between two and four weeks—closer to two weeks for orbits closer to the Lagrange point; the LL_3 family includes orbits with orbital periods of approximately four weeks [140]. These orbits are all unstable.

2.6.9.2 Distant Prograde Orbits Periodic three-body orbits certainly exist about the Earth and the Moon as well as the Lagrange points. Figure 2-23 illustrates the family of planar distant prograde orbits and shows how that family of orbits fits in between the family of L_1 and L_2 Lyapunov orbits. A spacecraft only needs to adjust its state slightly to transfer from a Lyapunov orbit to a distant prograde orbit and vice versa. This is explored in Section 2.6.11.3. Most distant prograde orbits are unstable; their orbital periods vary from two weeks to four weeks, much like the Lyapunov orbits.

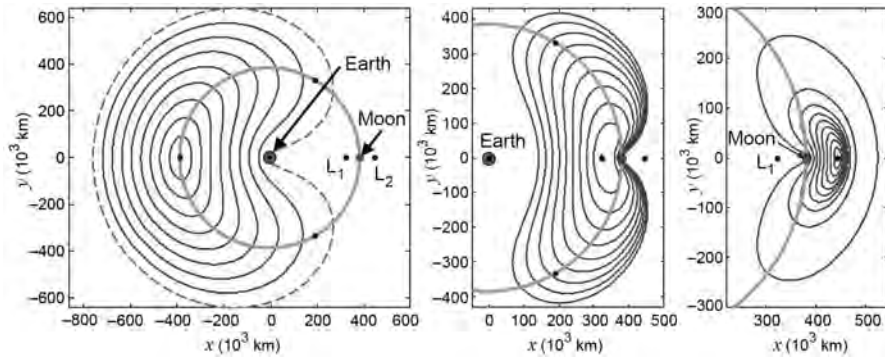


Figure 2-22 Example orbits in the families of Lyapunov orbits about the Earth–Moon L_3 point (left), L_1 point (middle), and L_2 point (right), viewed in the Earth–Moon rotating frame from above [46].

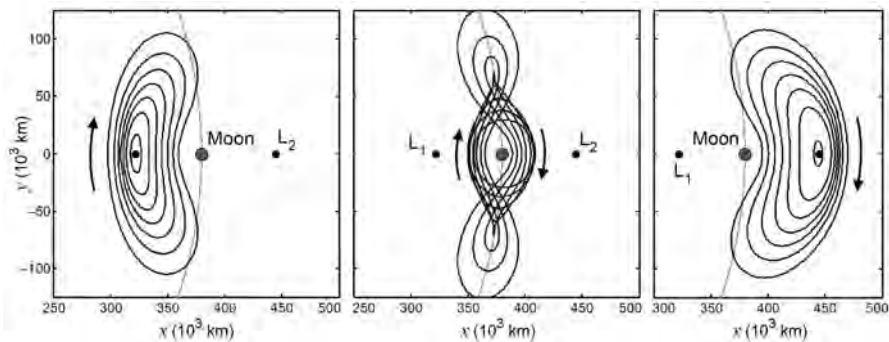


Figure 2-23 Example orbits in the families of Lyapunov orbits about the Earth–Moon L_1 point (left), distant prograde orbits about the Moon (middle), and Lyapunov orbits about the Earth–Moon L_2 point (right), viewed in the Earth–Moon rotating frame from above [46].

2.6.9.3 Distant Retrograde Orbits Distant retrograde orbits (DROs) are periodic three-body orbits that exist about the smaller primary, for example, the Moon in the Earth–Moon system, such that a spacecraft revolves about the body in a retrograde fashion. They are commonly found in the literature and in proposed spacecraft missions because they are frequently stable. They behave just like a normal two-body orbit, but occur in resonance with the motion of the three-body system. Figure 2-24 illustrates several examples of Earth–Moon DROs of varying radii from the Moon.

2.6.9.4 Halo Orbits Halo orbits are very well-known three-dimensional periodic solutions [119, 121, 122] to the circular restricted three-body problem. Fig-

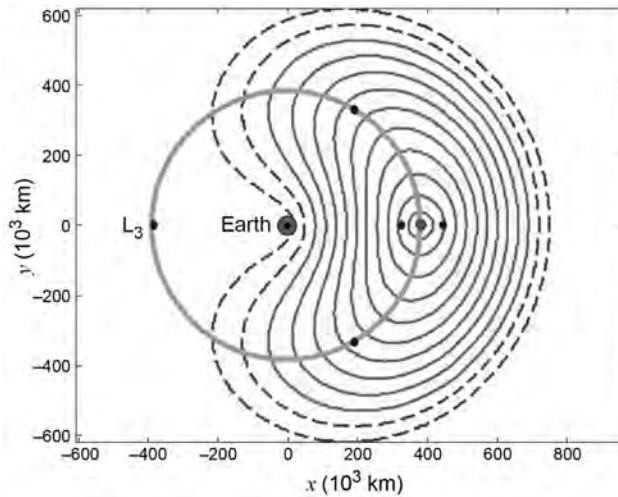


Figure 2-24 Example orbits in the family of distant retrograde orbits about the Moon, viewed in the Earth–Moon rotating frame from above.

Figure 2-25 shows a plot of several example halo orbits about the lunar L_1 and L_2 points. Many authors have studied how to take advantage of halo orbits for practical missions to the Moon [5–7]. Halo orbits are of particular use for lunar communication and navigation satellites [11]: a satellite in a halo orbit has an unimpeded view of both the Earth and either the near-side of the Moon or the far-side of the Moon, for lunar L_1 and L_2 halo orbits, respectively. Furthermore, a satellite may be placed in a halo orbit such that its view of the Sun is also never impeded, simplifying the satellite’s power and thermal systems.

Since the force field in the CRTBP is symmetric about the xy plane (see Section 2.5.1), and since halo orbits are asymmetric about this plane, each halo orbit solution to the CRTBP comes in a symmetric pair with a northern and a southern variety [121]. As one can see in Fig. 2-25, a satellite in a southern orbit spends more than half of its time below the Moon’s orbital plane, which gives that satellite benefits for communicating with objects in the southern hemisphere of the Moon.

It is convenient to specify a halo orbit by its z -axis amplitude, A_z , since one may formulate an analytical approximation to a halo orbit using that parameter as an input [106, 123, 124]. Other studies have specified a halo orbit using its Jacobi constant or its x_0 -value, namely, the x -value of the location where the orbit has a y -position of 0 km and a positive y -velocity in the synodic reference frame [46, 122]. Figure 2-26 shows several northern LL_1 and LL_2 halo orbits from the side in the synodic frame to illustrate the relationship between a halo orbit’s shape and its z -axis amplitude. Figure 2-27 shows the relationship between a halo orbit’s z -axis amplitude and its period for reference.

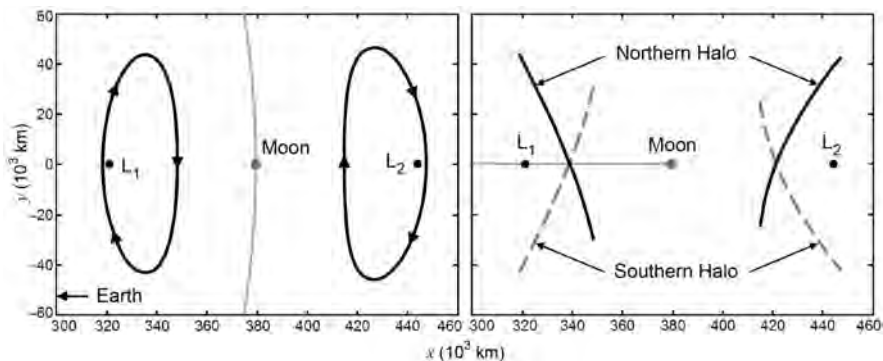


Figure 2-25 An illustration of four example halo orbits about the lunar L_1 and L_2 points. The halo orbits are viewed from above (left) and from the side (right) in the Earth–Moon synodic reference frame [47] (first published by the American Astronautical Society).

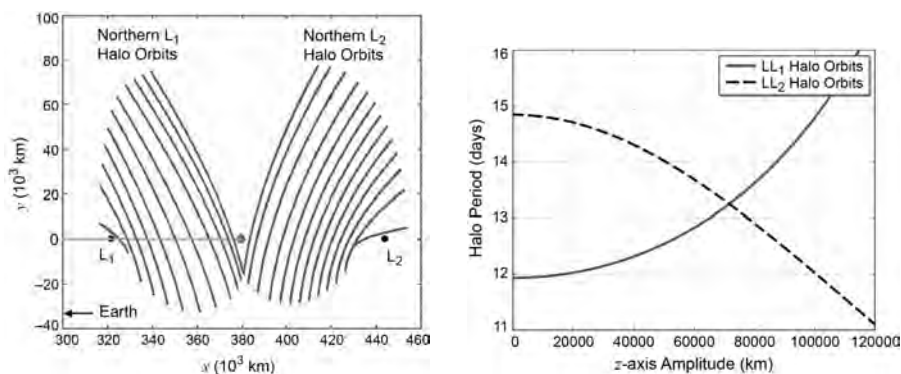


Figure 2-26 Northern Earth–Moon halo orbits [146] (Copyright © 2008 by American Astronautical Society Publications Office, San Diego, California [Web Site: <http://www.univelt.com>], all rights reserved; reprinted with permission of the AAS).

Figure 2-27 Earth–Moon halo orbit periods [44] (Copyright © 2009 by American Astronautical Society Publications Office, all rights reserved, reprinted with permission of the AAS).

2.6.9.5 Vertical Lyapunov Orbits Another family of libration orbits that exist about each of the collinear Lagrange points is the family of vertical Lyapunov orbits, also known as *vertical* orbits for short. Vertical orbits oscillate out of the xy plane, piercing the plane at the Lagrange point itself. They are symmetric orbits, traversing the same route above the plane as below it. Figure 2-28 provides several views of example orbits in the family of LL_1 vertical Lyapunov orbits.

2.6.9.6 Resonant Orbits Although there are numerous other interesting families of periodic orbits in the three-body system, the last type of orbit that will be

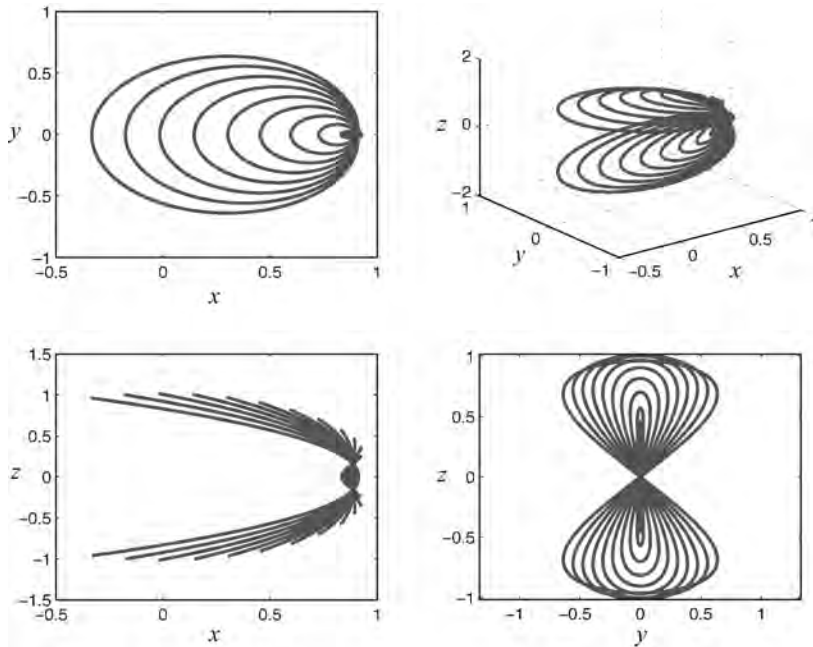


Figure 2-28 Four perspectives of example orbits within the family of L_1 vertical Lyapunov orbits computed in the Earth–Moon CRTBP.

described here is the resonant orbit. Resonant orbits in the Earth–Moon three-body problem are essentially two-body orbits about the Earth that are in resonance with the Moon, and which have been significantly perturbed by the Moon. As one may expect, there are different families of resonant orbits for each resonant period, namely, 3:1, 3:2, 5:1, 5:2, and so forth, where an $m:n$ resonant orbit is one where the spacecraft traverses the resonant orbit n times while the primaries orbit their barycenter m times. Figure 2-29 illustrates four families of resonant orbits in the Earth–Moon system, shown in the synodic reference frame.

The resonant orbits shown in Fig. 2-29 are particularly unstable as they pass by the Moon, but they are generally stable elsewhere. It is possible to transition a spacecraft off of one three-body orbit, such as a Lyapunov orbit, and onto a resonant orbit for very little fuel, if the transition is performed near the Moon. A spacecraft that arrives onto a resonant orbit may then sit in it without requiring any significant station-keeping fuel, until the spacecraft returns to the Moon. In that way, resonant orbits may play a useful role as a staging orbit, quarantine orbit, or a destination for a spacecraft to remain to avoid performing station-keeping maneuvers. One may also select how much time should pass between lunar swingbys, based on the resonance; for instance, a spacecraft traversing a 7:3 resonant orbit will spend far longer between lunar swingbys than a spacecraft traversing a 3:2 resonant orbit.

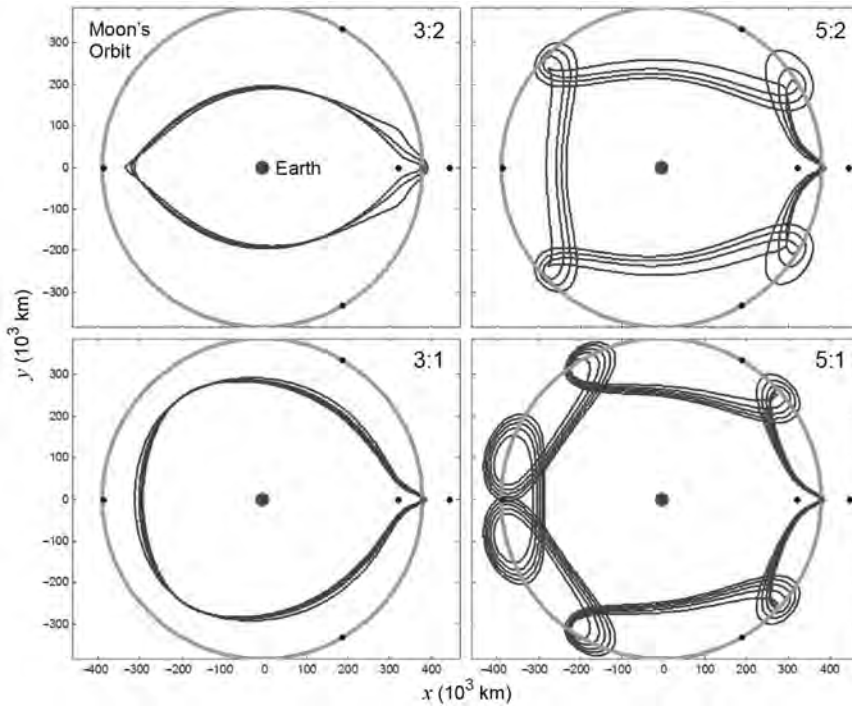


Figure 2-29 Four example families of resonant orbits in the Earth–Moon system, viewed from above in the Earth–Moon rotating frame.

2.6.10 Invariant Manifolds

The dynamics in the circular restricted three-body system permit the existence of five fixed points (Section 2.6.2.1) and numerous periodic orbits (Section 2.6.6.1). The three collinear libration points and many of the periodic orbit solutions in the Earth–Moon three-body system are unstable (Section 2.6.8). An unstable orbit has at least one stable and one unstable eigenvalue with corresponding eigenvectors. A spacecraft traveling along an unstable orbit that experiences a perturbation even slightly in the unstable direction will exponentially fall away from its nominal position on that orbit, tracing out a smooth trajectory away from the orbit. In a similar sense, a spacecraft that has the right initial conditions will follow a smooth trajectory that exponentially approaches an unstable orbit and eventually arrives on that orbit from the orbit’s stable direction. These two trajectories describe what is known as an orbit’s stable and unstable invariant manifolds.

An orbit’s *unstable invariant manifold* (W^U) contains the set of all trajectories that a spacecraft may take if it was perturbed anywhere on that orbit in the direction of the orbit’s unstable eigenvector. Similarly, an orbit’s *stable invariant manifold* (W^S) contains the set of all trajectories that a spacecraft may take to asymptotically

arrive onto that orbit along the orbit's local stable eigenvector. Put another way, the orbit's stable invariant manifold is the set of all trajectories that a spacecraft may take backward through time after a perturbation in the direction of the orbit's stable eigenvector.

Mathematically, the invariant manifolds are defined as follows. First, the CRTBP may be defined as a vector field bound in \mathbb{R}^6 . One and only one vector is bound to every point in the vector field. Thus, the integration of any point p in the vector field with respect to time generates only one trajectory. Let us define T_p as the trajectory generated by the point p . The α - and ω -limits are defined to be the set of points in \mathbb{R}^6 as T_p tends toward $-\infty$ and $+\infty$, respectively. The α - and ω -limits may include a single point, a periodic orbit, or, if T_p has no asymptotic behavior, they may include a large portion of the state space. The set of all points defining trajectories that have the same α -limit set is called the *unstable manifold* of that limit set. Similarly, the set of all points defining trajectories that have the same ω -limit set is called the *stable manifold* of that limit set.

2.6.10.1 Invariant Manifolds of the Unstable Lagrange Points The three collinear Lagrange points are unstable in both the Sun–Earth and Earth–Moon three-body systems; hence, they have associated invariant manifolds. Since the Lagrange points are single points in space, their invariant manifolds are one-dimensional structures. To produce them, one first computes the eigenvalues of the Jacobian of their states. If \mathbf{X} is the state of one of the collinear Lagrange points, equal to $[x \ y \ z \ \dot{x} \ \dot{y} \ \dot{z}]^T$, then its Jacobian is equal to

$$J = \frac{\partial \dot{\mathbf{X}}}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial y} & \cdots & \frac{\partial \dot{x}}{\partial z} \\ \frac{\partial \dot{y}}{\partial x} & \frac{\partial \dot{y}}{\partial y} & \cdots & \frac{\partial \dot{y}}{\partial z} \\ \frac{\partial \dot{z}}{\partial x} & \frac{\partial \dot{z}}{\partial y} & \cdots & \frac{\partial \dot{z}}{\partial z} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \ddot{x}}{\partial x} & \frac{\partial \ddot{x}}{\partial y} & \cdots & \frac{\partial \ddot{x}}{\partial z} \\ \frac{\partial \ddot{y}}{\partial x} & \frac{\partial \ddot{y}}{\partial y} & \cdots & \frac{\partial \ddot{y}}{\partial z} \\ \frac{\partial \ddot{z}}{\partial x} & \frac{\partial \ddot{z}}{\partial y} & \cdots & \frac{\partial \ddot{z}}{\partial z} \end{bmatrix} \quad (2.60)$$

After plugging in the equations of motion of the CRTBP given in Eqs. (2.1)–(2.3) in Section 2.5.1, Eq. (2.60) simplifies to

$$J = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\partial \ddot{x}}{\partial x} & \frac{\partial \ddot{x}}{\partial y} & \frac{\partial \ddot{x}}{\partial z} & 0 & 2 & 0 \\ \frac{\partial \ddot{y}}{\partial x} & \frac{\partial \ddot{y}}{\partial y} & \frac{\partial \ddot{y}}{\partial z} & -2 & 0 & 0 \\ \frac{\partial \ddot{z}}{\partial x} & \frac{\partial \ddot{z}}{\partial y} & \frac{\partial \ddot{z}}{\partial z} & 0 & 0 & 0 \end{bmatrix} \quad (2.61)$$

It is apparent that the Jacobian is the same as the A -matrix given in Eq. (2.16).

The eigenvalues of the Jacobian for each of the three collinear Lagrange points include two pairs of imaginary numbers and one pair of real numbers. Tables 2-6 and 2-7 summarize the six eigenvalues for the Jacobian of each of the five Lagrange points for the Earth–Moon system and for the Sun–Earth system, respectively. The eigenvector corresponding to the larger real eigenvalue indicates the unstable direction: \mathbf{v}^U ; the eigenvector corresponding to the other real eigenvalue indicates the stable direction: \mathbf{v}^S . The unstable manifold of the Lagrange point, W^U , may be mapped by propagating the state \mathbf{X}^U forward in time, where $\mathbf{X}^U = \mathbf{X} \pm \epsilon \mathbf{v}^U$ and ϵ is some small perturbation. Similarly, the stable manifold, W^S , may be mapped by propagating the state \mathbf{X}^S backward in time, where $\mathbf{X}^S = \mathbf{X} \pm \epsilon \mathbf{v}^S$.

The perturbation $\epsilon \mathbf{v}$ may be applied to the state \mathbf{X} in either a positive or a negative sense, corresponding to two halves of each manifold. One perturbation will result in motion that departs the Lagrange point toward the smaller body (for example, toward the Moon in the Earth–Moon system), and one will result in motion that departs the Lagrange point away from the smaller body. It is conventional to refer to the half of the manifold that moves toward the smaller body as the *interior* manifold, since it remains in the interior of the smaller body’s influence, at least for a short while, and

Table 2-6 A summary of the eigenvalues of the Jacobian of each Lagrange point in the Earth–Moon CRTBP.

Component	LL ₁	LL ₂	LL ₃	LL ₄	LL ₅
λ_1	-2.932056	-2.158674	-0.177875	1i	1i
λ_2	2.932056	2.158674	0.177875	-1i	-1i
λ_3	2.334386i	1.862646i	1.01041991i	0.95450078i	0.95450078i
λ_4	-2.334386i	-1.862646i	-1.01041991i	-0.95450078i	-0.95450078i
λ_5	2.268831i	1.786176i	1.00533144i	0.29820842i	0.29820842i
λ_6	-2.268831i	-1.786176i	-1.00533144i	-0.29820842i	-0.29820842i

Table 2-7 A summary of the eigenvalues of the Jacobian of each Lagrange point in the Sun–Earth CRTBP.

Component	EL ₁	EL ₂	EL ₃	EL ₄	EL ₅
λ_1	-2.532659	-2.484317	-0.002825	1i	1i
λ_2	2.532659	2.484317	0.002825	-1i	-1i
λ_3	2.0864535i	2.057014i	1.00000266i	0.99998974i	0.99998974i
λ_4	-2.0864535i	-2.057014i	-1.00000266i	-0.99998974i	-0.99998974i
λ_5	2.0152106i	1.985075i	1.00000133i	0.00453024i	0.00453024i
λ_6	-2.0152106i	-1.985075i	-1.00000133i	-0.00453024i	-0.00453024i

to refer to the half that moves away from the smaller body as the *exterior* manifold [37].

The process of analyzing and constructing the invariant manifolds of the unstable Lagrange points may be visualized by considering that each unstable Lagrange point is a dynamical saddle point, as illustrated by the plot shown in Fig. 2-30. One can see that a spacecraft's motion will follow the unstable manifold when propagated forward in time after a perturbation, and it will follow the point's stable manifold when propagated backward in time. Figure 2-30 also demonstrates how there are two halves of each manifold.

Figures 2-31–2-33 show plots of the stable and unstable manifolds of the first three Lagrange points in the Earth–Moon three-body system. The forbidden region is shown shaded in gray in each plot.

The eigenvalues of the Jacobian of the triangular Lagrange points include three imaginary pairs for the Sun–Earth and Earth–Moon three-body systems; hence, they do not have interesting associated invariant manifolds. A spacecraft following a trajectory near one of these Lagrange points will oscillate about the point. If the spacecraft is perturbed, its motion will change but it will not exponentially deviate from its nominal path.

2.6.10.2 Invariant Manifolds of Unstable Periodic Orbits Every unstable periodic orbit in the CRTBP has a set of invariant manifolds, much like the Lagrange points. The only substantial difference between the invariant manifolds of periodic

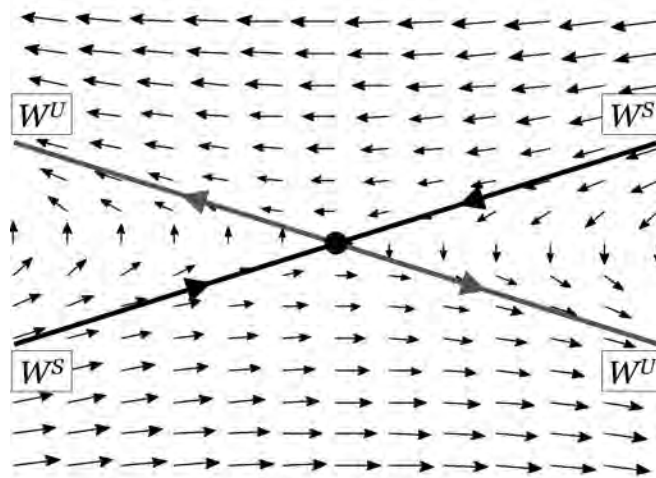


Figure 2-30 A dynamical saddle point, such as that of the unstable Lagrange points in the CRTBP, with a vector field shown that indicates the motion of a spacecraft near the point. There are two lines of stable (W^S) and unstable (W^U) manifolds of the saddle point (first published in Ref. [97]; reproduced with kind permission from Springer Science+Business Media B. V.).

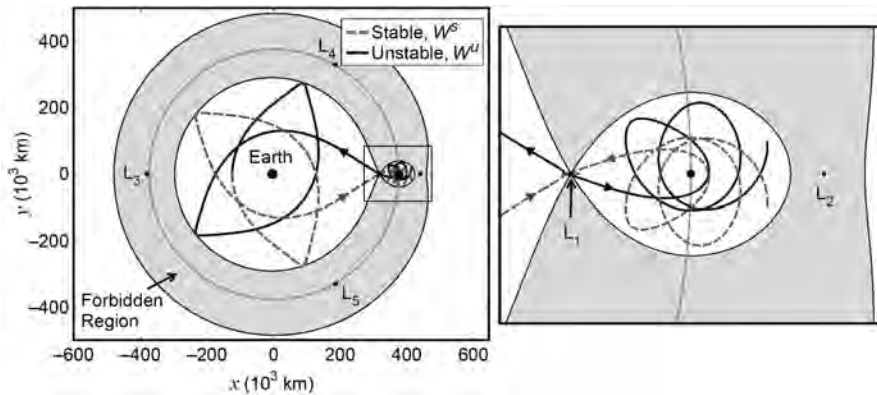


Figure 2-31 The stable and unstable invariant manifolds of the first Lagrange point of the Earth–Moon three-body system. *See inset at right for expanded view of the lunar vicinity* (First published in Ref. [97]; reproduced with kind permission from Springer Science+Business Media B. V.).

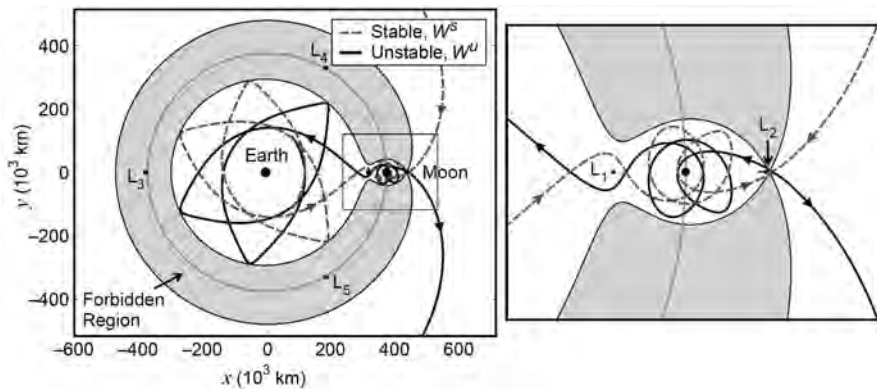


Figure 2-32 The stable and unstable invariant manifolds of the second Lagrange point of the Earth–Moon three-body system. *See inset at right for expanded view of the lunar vicinity* (First published in Ref. [97]; reproduced with kind permission from Springer Science+Business Media B. V.).

orbits and of the Lagrange points is that an additional dimension is added when considering periodic orbits: periodic orbits are one-dimensional structures where the Lagrange points are zero-dimensional structures. Consequently, the invariant manifolds of unstable periodic orbits are two-dimensional structures. They are constructed of a set of trajectories, where each trajectory corresponds to a point along

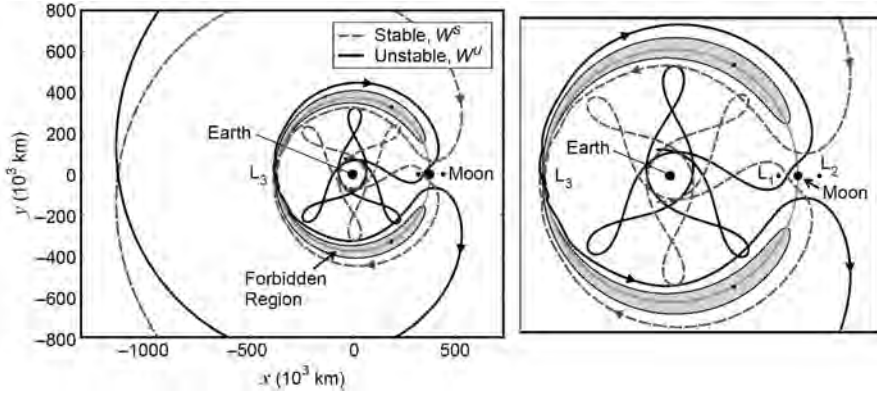


Figure 2-33 The stable and unstable invariant manifolds of the third Lagrange point of the Earth–Moon three-body system. *See inset at right for an expanded view in the vicinity of Earth* (First published in Ref. [97]; reproduced with kind permission from Springer Science+Business Media B. V.).

the periodic orbit. The set of trajectories wraps about itself, forming a topological tube. This is further explained below.

To produce the invariant manifolds of an unstable periodic orbit, one requires information about the local stability characteristics of each point along the orbit. In theory, one may evaluate the eigenvalues and eigenvectors of the Jacobian at each and every state along the orbit, and use that information to produce the orbit's invariant manifold. However, evaluating so many eigenvalues requires a great deal of computation. A more efficient manner of producing the invariant manifolds uses the eigenvalues and eigenvectors of the monodromy matrix [147, 148].

Since the monodromy matrix is produced by propagating the state transition matrix all the way around the orbit, from time $t = t_0$ to time $t = t_0 + T$, it contains information about the stability of the entire orbit. To determine the stable and unstable directions at each point along the orbit, one only has to propagate the stable and unstable eigenvectors of the monodromy matrix about the orbit using the state transition matrix. That is, the stable and unstable vectors at time t_i about the orbit, \mathbf{v}_i^S and \mathbf{v}_i^U , respectively, may be determined using the stable and unstable eigenvectors of the monodromy matrix, \mathbf{v}^S and \mathbf{v}^U , respectively, using the following equations

$$\mathbf{v}_i^S = \Phi(t_i, t_0) \mathbf{v}^S \quad (2.62)$$

$$\mathbf{v}_i^U = \Phi(t_i, t_0) \mathbf{v}^U \quad (2.63)$$

A small perturbation, ϵ , is then applied to the state of the orbit at that time, \mathbf{X}_i , and the result is propagated in time. Since the state transition matrix grows exponentially along an unstable orbit, the magnitudes of the vectors \mathbf{v}_i^S and \mathbf{v}_i^U grow along the orbit. It is therefore important to normalize the vectors so that a consistent perturbation is

applied to each orbit state. The final equations to produce the initial conditions for the stable and unstable manifolds at time t_i about the orbit, \mathbf{X}_i^S and \mathbf{X}_i^U , respectively, are then equal to

$$\mathbf{X}_i^S = \mathbf{X}_i \pm \epsilon \frac{\mathbf{v}_i^S}{|\mathbf{v}_i^S|} \quad (2.64)$$

$$\mathbf{X}_i^U = \mathbf{X}_i \pm \epsilon \frac{\mathbf{v}_i^U}{|\mathbf{v}_i^U|} \quad (2.65)$$

The sign of the perturbation differentiates between interior and exterior manifolds, as discussed in Section 2.6.10.1.

Some discussion should be provided regarding the magnitude of the perturbations applied to the state to produce the manifolds. The theoretical invariant manifolds of the orbit include the set of all trajectories that asymptotically approach the orbit as time goes either forward or backward. In fact, they never truly arrive on the orbit in finite time, but just come arbitrarily close to the orbit. To map them, one approximates the manifolds by perturbing a state slightly off of the orbit and then propagating that state in time. The smaller the perturbation, the closer the approximation comes to mapping the true manifolds; however, small perturbations require more time to depart from the orbit than larger perturbations. When designing practical missions, one is less interested in precisely mapping the invariant manifolds of the orbits, and generally more interested in computationally-swift algorithms. Additionally, the dynamics of the trajectories depend the greatest on the largest eigenvalues since motion in those directions grows exponentially faster than motion in any other direction. Hence, somewhat large perturbations may be used to map out the motion of the trajectories in the manifolds, for example, on the order of 100 km in the Earth–Moon system and 1000 km in the Sun–Earth system. In practice, the perturbation magnitudes are given in either units of position or units of velocity, but the perturbation is applied proportionally to all six components. A 100-km perturbation means that the magnitude of the perturbation applied to the position coordinates is equal to 100 km, and the resulting proportionality is used to apply the perturbation to the velocity components, that is

$$\epsilon = \frac{100 \text{ km}}{\sqrt{v_x^2 + v_y^2 + v_z^2}} \quad (2.66)$$

The structure of the manifolds of an orbit greatly depends on the stability characteristics of each portion of the orbit. Orbits such as libration orbits are fairly uniformly unstable; that is, the local Lyapunov exponent does not vary much along the orbit (Anderson, among others, provides a detailed exploration about the local Lyapunov exponent of libration orbits [149]). Consequently, their manifolds are fairly smooth as they extend from the orbit. Various other orbits are unstable due to a close flyby of one of the primary bodies. The local stability of these orbits changes drastically, becoming very unstable as the orbit approaches one of the massive bodies. Hence, their manifolds spread out quickly near the body and remain fairly close together elsewhere.

Figure 2-34 shows the stable and unstable manifolds of a two-dimensional Lyapunov orbit about the Earth–Moon L_2 point. One can see that the manifolds are smooth and form a tube-like structure. They remain well-defined until they encounter the Moon, at which time they spread out very rapidly, and the tube-like structure becomes less obvious. One can also see that the stable and unstable manifolds are

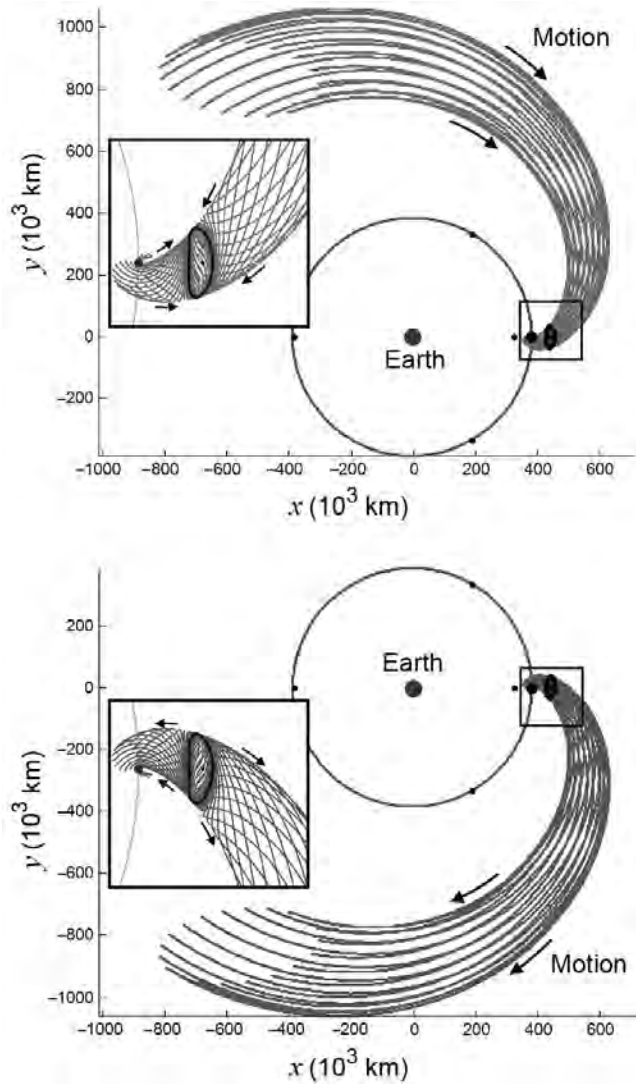


Figure 2-34 The stable (left) and unstable (right) manifolds of a Lyapunov orbit about the Earth–Moon L_2 point.

symmetric about the x -axis, due to the symmetry in the CRTBP. For comparison, Fig. 2-35 shows the stable manifold of a resonant flyby orbit, which shows how the structure of the manifolds depends on the local stability of the orbit. One notices that the trajectories in the manifold diverge quickly near the Moon but remain near the orbit elsewhere. The unstable manifold is not shown but is symmetric to the stable manifold.

2.6.10.3 Invariant Manifolds of Unstable Quasiperiodic Orbits Unstable quasi-periodic orbits have associated stable and unstable invariant manifolds, much like unstable periodic orbits; however, the structure and the procedures required to produce them are slightly different. Quasiperiodic Lissajous and quasi-halo orbits in the CRTBP are two-dimensional structures [125]. Hence, their invariant manifolds are three-dimensional structures. The additional dimension adds benefits as well as complexity when using them in practical mission designs.

Since quasiperiodic orbits never retrace their path, one cannot produce them entirely, although one can use a variety of numerical tools to represent them and to produce desirable segments of them [150]. Since these orbits are not periodic, they do not have associated monodromy matrices. Hence, one cannot use the same simplified procedures to produce their invariant manifolds as those procedures discussed in Section 2.6.10.2.

To produce a quasiperiodic orbit's invariant manifolds, one can always compute the eigenvectors of the Jacobian of the states at each point along sample segments of the orbit, and follow the same procedures as given in the previous sections. However, that procedure is numerically intensive and slow. Alternatively, to reduce the computational load, one may approximate the manifolds by producing an analog to the monodromy matrix. One may propagate the state transition matrix from one $y = 0$ plane crossing in the synodic frame to the next (or to any later crossing) and

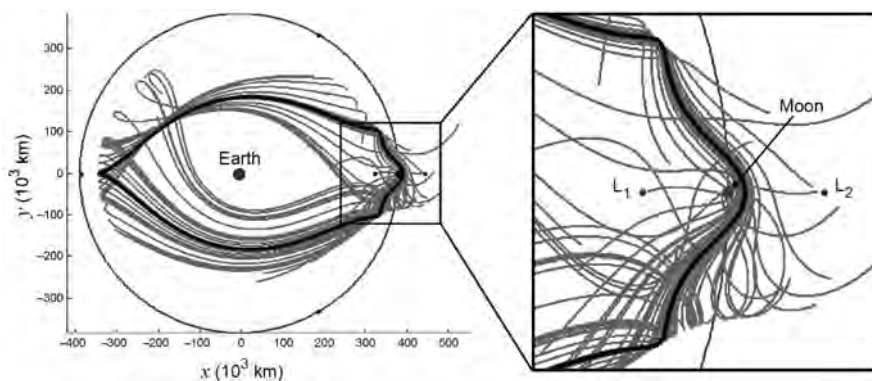


Figure 2-35 The stable manifold of a resonant lunar flyby orbit. See inset at right for expanded view of the lunar vicinity.

use the resulting matrix as a pseudo-monodromy matrix. When one propagates this matrix's stable and unstable eigenvectors about the orbit segment, and then follows the process outlined in Section 2.6.10.2, one produces approximations of the invariant manifolds of the quasiperiodic orbit. These approximations are often good enough to be used for preliminary spacecraft mission design, such as that used for the *Genesis* spacecraft mission [137].

2.6.11 Orbit Transfers

Dynamical systems theory provides the tools needed to systematically produce transfers to/from unstable orbits in the CRTBP. This section discusses several example orbit transfers as demonstrations of the application of dynamical systems theory. Section 2.6.11.1 discusses the construction of a transfer from the Earth to a halo orbit about the Sun–Earth L_2 point, and that transfer is very similar to that used by the *WMAP* mission [70]. Section 2.6.11.3 discusses the construction of a chain of periodic orbits in the CRTBP, which is relevant to missions like *Genesis* [71, 72] and *Wind* [63]. That is, trajectories are constructed that transfer a spacecraft back and forth between several periodic orbits in the CRTBP. These examples demonstrate the procedures that may be followed to construct any type of orbit transfer in the CRTBP using dynamical systems theory.

2.6.11.1 Surface to Orbit Transfers Several missions (including *WMAP*, *Herschel*, and *Planck*) have demonstrated the benefits of operating in a libration orbit about the Sun–Earth L_2 point; many other missions have been proposed to operate in similar orbits, including the *James Webb Space Telescope* and the *Terrestrial Planet Finder*. In this section, we demonstrate how to construct a ballistic transfer from the Earth to a halo orbit about EL_2 , a transfer that might prove to be very useful for missions such as these proposed missions. The transfers produced here do not require any orbit insertion maneuvers; after their LEO departures, each transfer is thereafter entirely free of any deterministic maneuvers. The process used here may be generalized to compute a transfer from the surface of the secondary body in most three-body systems into many unstable three-body periodic orbits, or vice-versa.

We first consider the family of halo orbits about the EL_2 point, illustrated in Section 2.6.9.4. The family begins as a bifurcation of the family of planar Lyapunov orbits about EL_2 . The orbits in the family gradually move farther out of the plane until they eventually make close approaches with the Earth. Example orbits in the family of northern EL_2 halo orbits are shown in Fig. 2-36.

We next consider a single unstable halo orbit and produce its stable invariant manifold. This manifold includes all the trajectories that a spacecraft may take to arrive onto the orbit. A plot of the example halo orbit and its stable manifold is shown in Fig. 2-37. The trajectories shown in blue have a perigee altitude below 500 km. The halo orbit chosen here has a Jacobi constant equal to approximately 3.00077207. The CRTBP is a good model of the real Solar System for trajectories propagated for a reasonably short amount of time, namely, for one orbital period of the two primary masses about their barycenter, or about a year in the Sun–Earth system. Beyond

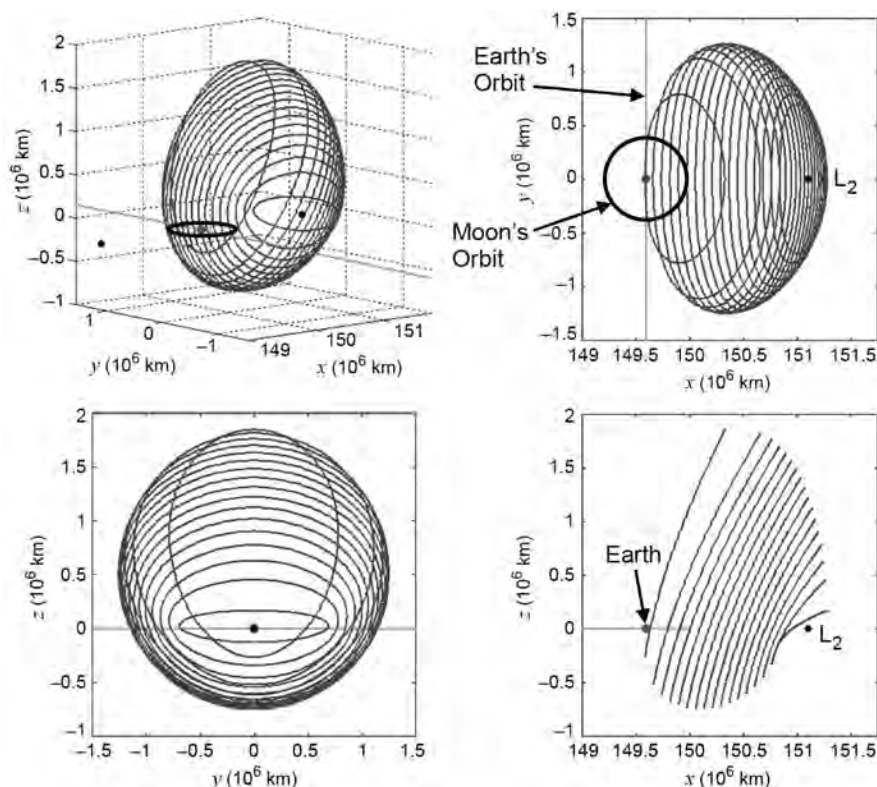


Figure 2-36 Example halo orbits in the family of northern halo orbits about the Sun–Earth L_2 point. The orbits are shown from four perspectives.

that, the accumulation of errors due to perturbations in the real Solar System causes the CRTBP approximation to break down. The trajectories shown in Fig. 2-37 have been propagated for at most 365 days—they are only plotted in the figure until they cross the plane of the Earth for clarity. As the propagation time is increased, the trajectories may make additional close approaches to the Earth. In some cases the second or third perigee passes closer to the Earth than the first. These features will be explored below.

Each trajectory shown in Fig. 2-37 may be characterized using several parameters. The parameter τ , defined in Section 2.6.2.3, indicates the point where the trajectory arrives at the halo orbit. The closest approach of each trajectory with the Earth is identified to compute the perigee altitude and ecliptic inclination with respect to the Earth. These two parameters are useful because they indicate the altitude and inclination of a low Earth orbit that may be used as a staging orbit to transfer a spacecraft to the halo orbit. Figure 2-38 shows two plots: one of the perigee altitude and one of the corresponding ecliptic inclination as functions of τ , where the vertical

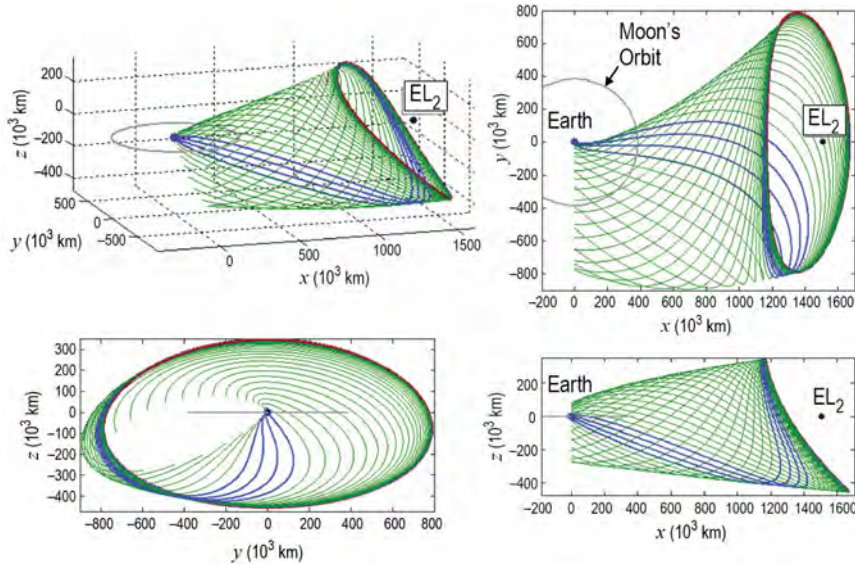


Figure 2-37 An example unstable halo orbit (green) about the Sun–Earth L_2 point and its stable invariant manifold (blue). (See insert for color representation of this figure.)

bars indicate the locations in the manifold that have perigee altitudes below 500 km. For example, one can see that the trajectory with a τ -value of 0.751 encounters a closest approach with the Earth with a perigee altitude of approximately 185 km and an ecliptic inclination of approximately 34.8 deg. Hence, a spacecraft in a circular low Earth parking orbit with an altitude of 185 km and an ecliptic inclination of 34.8 deg may perform a tangential ΔV to transfer onto the manifold; once on the manifold, the spacecraft ballistically follows it and asymptotically arrives on the halo orbit.

There are two statements in these results that need to be addressed. The first is that the inclination values displayed in Fig. 2-38 are the inclination values computed in the axes of the CRTBP: namely, in a plane that is very similar to the ecliptic. The equatorial inclination values of these perigee points depend on which date a spacecraft launches. Since the Earth's rotational axis is tilted by approximately 23.45 deg with respect to the ecliptic [97], many equatorial inclinations may be used to inject onto a desired trajectory, depending on the date. The second statement that should be addressed is that the results shown in Fig. 2-38 depend greatly on the perturbation magnitude, ϵ , described in Section 2.6.10. Implementing a different perturbation magnitude results in a change in the τ -values required to obtain a certain trajectory. For example, if ϵ were reduced, the trajectories modeling the orbit's manifold would spend more time asymptotically approaching/departing the orbit. Once the trajectories are sufficiently far from the orbit, their characteristics are nearly

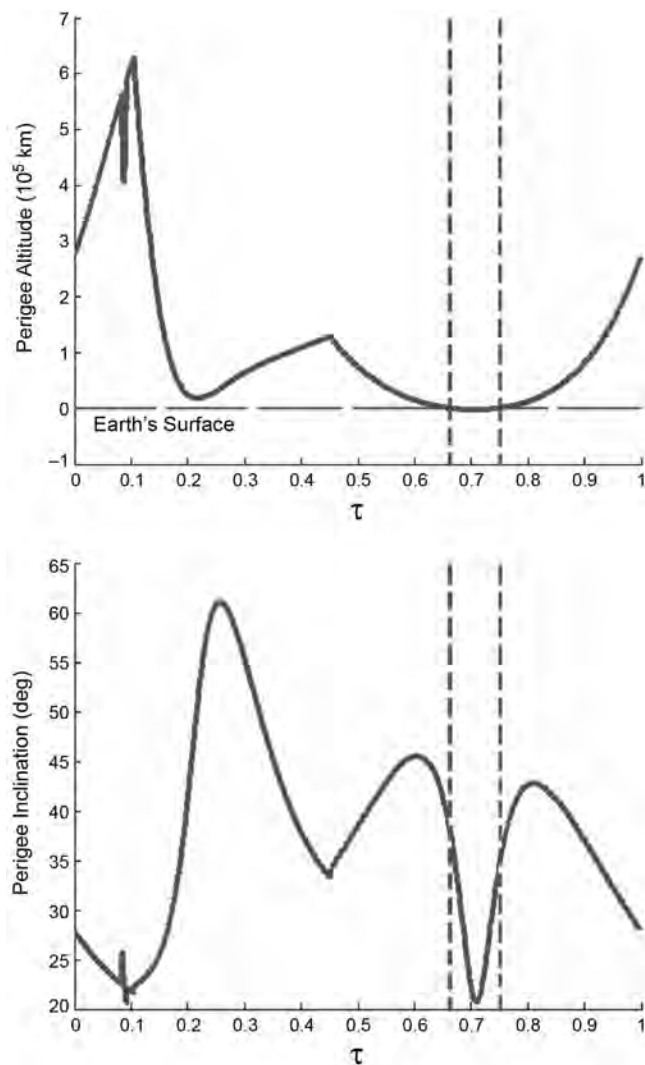


Figure 2-38 The perigee altitude (top) and corresponding ecliptic inclination (bottom) of the trajectories in the stable manifold shown in Fig. 2-37 as functions of τ .

unchanged. The result is that the τ -value for a desired trajectory is strongly related to the value of ϵ . This has no significant effect for practical spacecraft mission designs; a spacecraft following a trajectory in the halo orbit's stable manifold will asymptotically approach the halo orbit—the value of ϵ is only used for modeling the stable manifold.

It is now of interest to identify how the manifolds change and how the plots shown in Fig. 2-38 change as the manifold's propagation time is varied. Figure 2-39 shows plots of the stable manifold of the same halo orbit propagated for successively longer amounts of time. One can see that the trajectories on the manifold spend some amount of time near the halo orbit (where, once again, the amount of time depends on the value of ϵ), and then depart. It may be seen that many of the trajectories in the manifold make closer approaches with the Earth after their first perigee. Figure 2-40 shows many plots of the closest approach each manifold makes with the Earth for varying amounts of propagation time. It is clear that the longer propagation times yield closer perigee passages.

The procedures given in this section may be repeated for each halo orbit in the entire family of halo orbits, and maps may be produced showing the range of perigee altitudes and the range of inclination values obtainable for each halo orbit. These are useful for identifying the approximate location of desirable trajectories in the real Solar System.

2.6.11.2 Homoclinic and Heteroclinic Connections Many unstable periodic orbits in the CRTBP contain *homoclinic* connections with themselves and/or *heteroclinic* connections with other unstable periodic orbits [71, 98, 151]. If a trajectory in an orbit's unstable manifold departs that orbit, traverses the three-body system for some time, and then later arrives back onto the same orbit, it makes what is known as a homoclinic connection with the host orbit [151]. This trajectory is contained within both the orbit's unstable and stable manifolds. McGehee proved the existence of homoclinic connections in both the interior and exterior regions of the three-body system [152]. In a similar sense, a different trajectory within the unstable manifold of one orbit may depart that orbit and eventually arrive onto a second orbit. The trajectory is thus contained within both the unstable manifold of the first orbit and the stable manifold of the second orbit. Such a trajectory forms what is known as a heteroclinic connection between the two unstable orbits [151].

In theory, heteroclinic connections asymptotically depart one orbit and asymptotically approach another orbit. In practice, the spacecraft is never truly on any host periodic orbit, but is instead within some small distance from the orbit. For the purpose of the discussions provided here, it is assumed that a spacecraft departs an orbit when its state is perturbed off of that orbit, and it arrives on the new orbit when it arrives at the state that corresponds to the perturbation that generated the stable manifold. For the case of orbit transfers in the Earth–Moon system, this means that the duration of an orbit transfer includes all time that the spacecraft is further than 100 km from a host orbit.

Many authors have explored homoclinic and heteroclinic transfers between three-body orbits as transport mechanisms for spacecraft and comets [98, 147, 149, 151, 153–160]. Using dynamical systems theory, Lo and Ross noted that the orbit of the comet Oterma appeared to shadow the invariant manifolds of libration orbits about the L_1 and L_2 points in the Sun–Jupiter three-body system [161]. Koon et al. later showed that the comet closely followed a homoclinic-heteroclinic chain [151]. Gómez et al. began exploring the numerical construction of orbits with prescribed

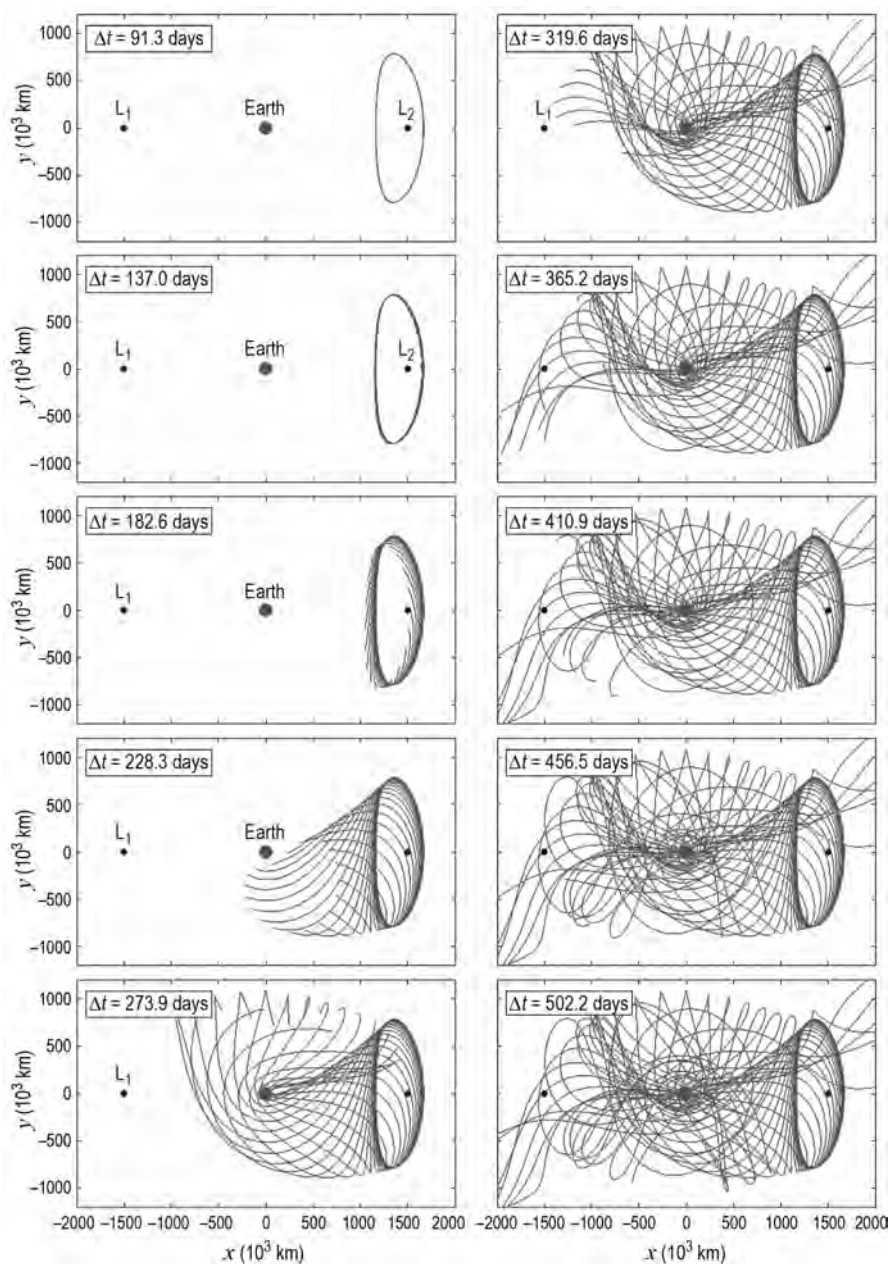


Figure 2-39 The stable manifold of a Sun–Earth L_2 halo orbit propagated for successively longer amounts of time. The duration of each propagation is indicated in each plot by the value Δt .

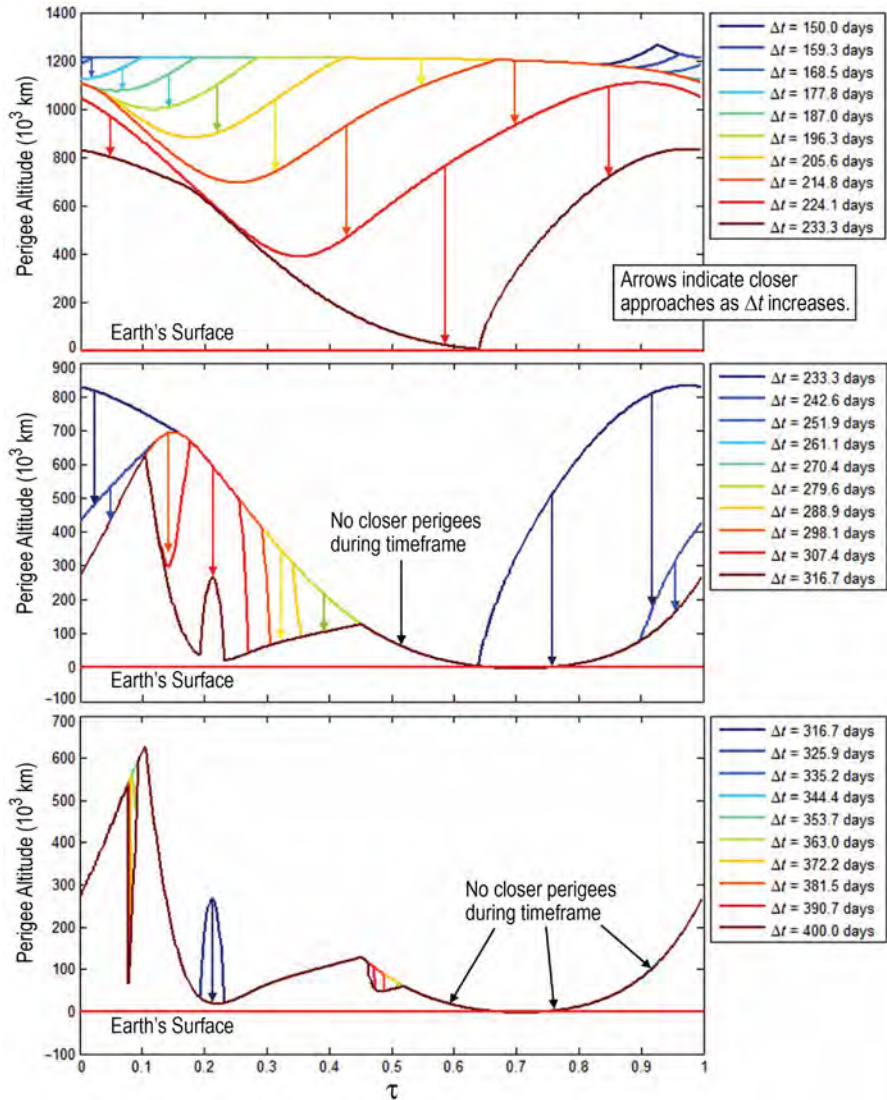


Figure 2-40 The altitude of closest approach of each trajectory in the stable manifolds shown in Fig. 2-39 with respect to the Earth. The propagation times of each manifold are shown in the legend. Longer propagation times yield closer perigee passages. (See insert for color representation of this figure.)

itineraries to describe the resonant transitions exhibited by the comet Oterma [98]. The material presented in Section 2.6.11.3 extends their work, applying a method for the construction of prescribed orbit transfers in the Earth–Moon system [101].

2.6.11.3 Orbit Transfers and Chains Once a spacecraft is on an unstable periodic orbit in the three-body system, then it may theoretically stay there for an arbitrarily long time, or it may depart that orbit by following any trajectory on that orbit’s unstable manifold. The practical ΔV cost for a spacecraft to depart an orbit is the same as the cost of station-keeping to remain on that orbit: both are arbitrarily small given good navigational support. These considerations are further explored in Chapter 6.

Section 2.6.10.2 shows several examples of stable and unstable invariant manifolds of unstable periodic orbits. One may notice by studying these manifolds that by controlling exactly when the spacecraft departs from its periodic orbit, it may be able to transfer to numerous other locations in the state space, including, but not limited to, the surface of the Moon, any of the five lunar Lagrange points, another unstable periodic orbit in the system, or an escape trajectory away from the vicinity of the Moon or Earth. If the spacecraft were carefully navigated onto the correct trajectory within the unstable manifold of one orbit, it would then encounter the stable manifold of a different unstable three-body orbit.

After considering a spacecraft’s options, several categories of orbit transfers may be identified. Table 2-8 summarizes a few characteristic categories of orbit transfers. In the table, a “stable orbit” includes conventional two-body orbits about either of the two primaries in the system, as well as stable three-body orbits, and even transfers to/from the surface of one of the primary bodies. The minimum number of ΔV s indicates the fewest number of maneuvers that may typically be used to perform the given transfer. There are many cases when a particular transfer might require more maneuvers, such as a transfer from the surface of a body to a particular orbit in space with a time constraint. There are also certain special cases when a transfer might require fewer maneuvers, such as a transfer between two stable orbits where the two orbits intersect in space. Nonetheless, Table 2-8 gives a good idea about the minimum number of required maneuvers for orbit transfers in several circumstances.

Table 2-8 A summary of several categories of orbit transfers in the CRTBP.

Orbit 1	Orbit 2	Constraints	Minimum Number of ΔV s
Stable Orbit	Stable Orbit	None	2
Stable Orbit	Unstable Orbit	None	1
Unstable Orbit	Stable Orbit	None	1
Unstable Orbit	Unstable Orbit	None	1
Any Orbit	Any Orbit	Transfer Time	2
Unstable Orbit	Unstable Orbit	Same Jacobi Constant	0

In this section, low-energy orbit transfers are introduced that may be useful for practical mission design, and which are useful background for the discussions of low-energy lunar transfers in later chapters. These orbit transfers are in the category represented by the last row of Table 2-8, transfers that are free of deterministic maneuvers.

Low-energy transfers between unstable orbits may be located in the CRTBP by analyzing Poincaré maps (Section 2.6.3) [156]. Suppose there are two unstable Lyapunov orbits in the Earth–Moon three-body system: one about LL_1 and the other about LL_2 . Both of these orbits have a set of stable and unstable invariant manifolds. In the planar CRTBP, each point along a manifold may be characterized by a four-dimensional state $[x, y, \dot{x}, \dot{y}]$. If a surface of section is placed in \mathbb{R}^4 at some x -position, the resulting intersection is a surface in \mathbb{R}^3 . If it is further specified that the two Lyapunov orbits have the same Jacobi constant, then each point along any trajectory within both orbits' manifolds will have the same Jacobi constant and the phase space of the problem is reduced to \mathbb{R}^2 . The state at any intersection in the surface may only be reconstructed if the Poincaré map is one-sided, since the Jacobi constant has a sign ambiguity. The stable and unstable manifolds of both orbits appear as curves in the two-dimensional Poincaré map. Any intersection of these curves corresponds to a free transfer between the two orbits.

Figure 2-41 illustrates the process of identifying free transfers from a Lyapunov orbit about LL_1 to a Lyapunov orbit about LL_2 . In this case, the value of the Jacobi constant of both orbits has been selected to be 3.13443929. A P_+ Poincaré map has been constructed, where the surface, Σ , has been placed at the x -coordinate of the Moon, namely, at a value of approximately 379,730 km with respect to the barycenter of the Earth–Moon system. The top of Fig. 2-41 illustrates the unstable and stable manifolds integrated to the first intersection with the surface of section. The intersection of both manifolds with the surface of section is shown on the bottom of Fig. 2-41. One can see that there are two intersections that correspond to the two free transfers indicated in the figure.

The simple illustration shown in Fig. 2-41 may be extended by propagating the manifolds longer and identifying intersections in the manifolds that correspond to longer, more complicated heteroclinic connections. The Poincaré map shown in Fig. 2-42 is produced by propagating the unstable manifold of the LL_1 Lyapunov orbit and the stable manifold of the LL_2 Lyapunov orbit for 60 days each. In addition, the map shown in Fig. 2-42 is a P_{\pm} map, displaying all intersections of both manifolds with the surface of section. In this particular mapping, the majority of the points shown below the $y = 0$ line are members of the P_+ map (including the points shown in Fig. 2-41), the majority of the points shown above it are members of the P_- map, and all observed intersections of the two manifolds do indeed intersect, even accounting for the sign ambiguity of \dot{x} .

Figure 2-42 includes eight example orbit transfers to illustrate what sort of heteroclinic connections exist between these two libration orbits. Certain types of motion appear in more than one heteroclinic connection. For example, the trajectories labeled (1), (2), and (7) appear to graze a distant prograde orbit, whereas the trajectories labeled (1), (3), and (4) appear to traverse a figure-eight type orbit. The appearance of

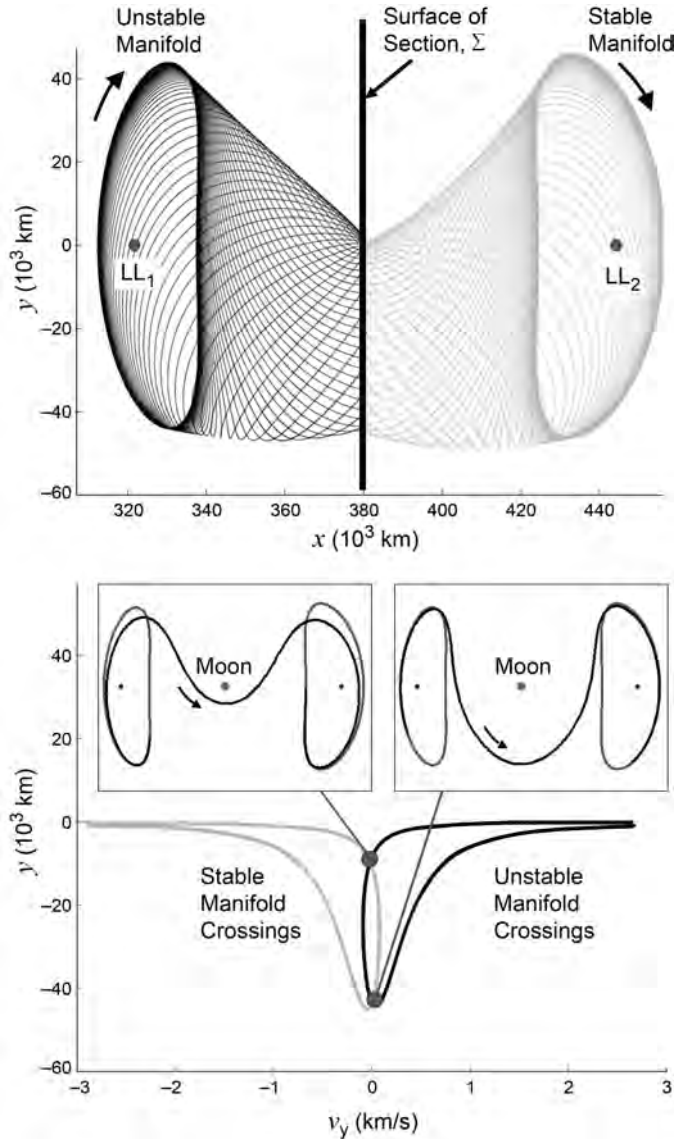


Figure 2-41 An illustration of the process of using a Poincaré map to identify free transfers between two Lyapunov orbits. Both orbits have a Jacobi constant of 3.13443929. Top: the unstable manifold of an LL_1 Lyapunov orbit and the stable manifold of an LL_2 Lyapunov integrated to the surface of section. Bottom: the corresponding P_+ Poincaré map and two free transfers [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

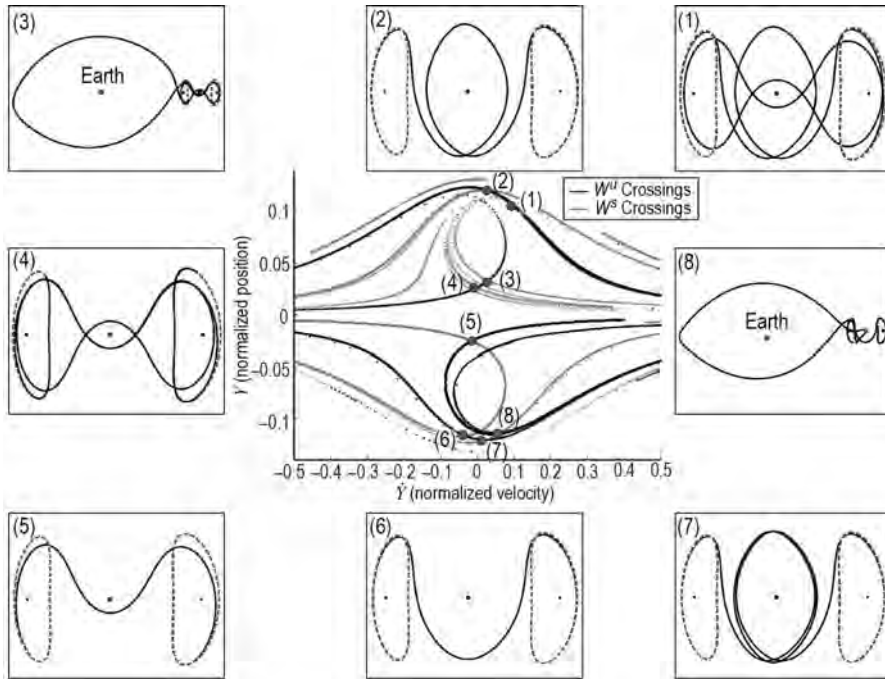


Figure 2-42 The P_{\pm} Poincaré map produced from the same system and surface of section shown in Fig. 2-41, now with an extended manifold propagation duration of 60 days. The plots shown surrounding the Poincaré map illustrate several example free transfers that have been identified in the central map [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

such orbits in the Poincaré maps reinforces the idea that one may construct a specific chain of simple orbits to construct a complicated itinerary of orbit transfers.

The Poincaré map is a useful tool to identify what sorts of orbit transfers exist, but it does not immediately reveal the shape or geometry of the transfers. For instance, the transfer labeled (8) in Fig. 2-42 includes a lunar flyby, which may or may not be desirable. Section 2.6.12 introduces a method that may be used to construct a desirable sequence of orbit transfers after identifying that such orbit transfers exist.

Free transfers only exist in the CRTBP between two unstable orbits that have the same Jacobi constant. Figure 2-43 shows a plot of several families of three-body orbits in the Earth–Moon CRTBP, where the orbits’ Jacobi constant values are plotted as functions of their x_0 -values. The curves shaded in black correspond to unstable three-body orbits; the curves shaded in gray correspond to orbits that are neutrally stable [130]. The horizontal line indicates the Jacobi constant value used to produce

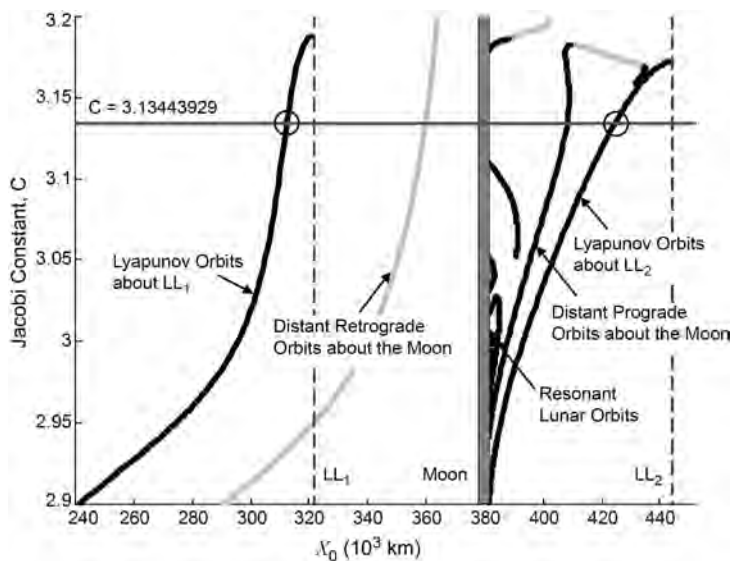


Figure 2-43 A plot of several families of three-body orbits in the Earth–Moon CRTBP, where the orbits’ Jacobi constant values are plotted as functions of their x_0 -values. The curves shaded in black correspond to unstable three-body orbits; the curves shaded in gray correspond to orbits that are neutrally stable [162] (Copyright ©2006 by American Astronautical Society Publications Office (web site: <http://www.univelt.com>), all rights reserved, reprinted with permission of the AAS).

the heteroclinic connections observed in Figs. 2-41 and 2-42. The figure verifies that the families of Lyapunov orbits about LL_1 and LL_2 both include unstable orbits at the same indicated Jacobi constant value, along with the family of distant prograde orbits, which helps to explain the appearance of such an orbit in the transfers labeled (1), (2), and (7) in Fig. 2-42.

2.6.12 Building Complex Orbit Chains

In the previous section, a technique was presented that may be used to identify the heteroclinic connections between two unstable periodic orbits. Previous papers have theorized using symbolic dynamics that if a heteroclinic connection exists, it is possible to find a trajectory that transfers back and forth arbitrarily between those orbits. Robinson provides a thorough review of the background of symbolic dynamics [163]. Canalias et al. [155], provide a methodology to search for a combination of homoclinic transfers that may be used to change the phase of a spacecraft traversing an unstable periodic orbit. In this section we study a practical method to construct a complex orbit chain given a desired sequence of homoclinic and/or heteroclinic transfers.

2.6.12.1 Constructing a Complex Orbit Chain One may describe a spacecraft's itinerary between simple periodic orbits in the CRTBP by considering its state at each x -axis crossing. A spacecraft traversing any simple periodic orbit pierces the x -axis twice: once with positive and once with negative values of \dot{y} . One may model a simplified orbit transfer by considering that the spacecraft departs the initial orbit at one x -axis crossing, is midway through the transfer at the next x -axis crossing, and completes the transfer at a later x -axis crossing. Using this conceptualization, one may construct a set of x -axis states to describe a given itinerary between two orbits. A set of eight states are summarized in Fig. 2-44 and Table 2-9 for transfers between an example LL_1 Lyapunov orbit and an example distant prograde orbit (DPO) that have the same Jacobi constant.

The states given in Fig. 2-44 and Table 2-9 have been collected from two sources. The states corresponding to the simple periodic orbits (A, B, E, and F) have been taken directly from those orbits; the algorithm described in Section 2.6.6.2 is well-suited to generate the states of a periodic orbit at their orthogonal x -axis crossings. The states that correspond to the orbit transfers (C, D, G, and H) have been taken from their heteroclinic connections identified using the Poincaré analysis described in Section 2.6.3. A theoretical heteroclinic connection between these orbits asymp-

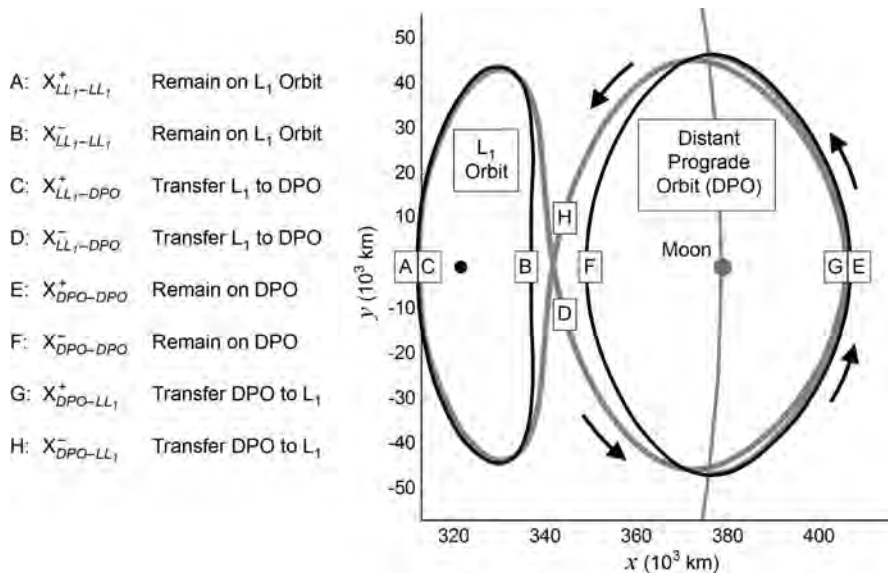


Figure 2-44 A summary of the states needed to produce complex itineraries between two orbits. In this case, the two orbits are a Lyapunov orbit about L_1 and a DPO about the Moon. The states “D” and “H” are on the x -axis, although the labels are offset [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

Table 2-9 The eight states shown in Fig. 2-44. The state coordinates are given in the Earth–Moon synodic reference frame, relative to the Earth–Moon barycenter, in both nondimensional normalized units and SI units [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

	State	Units	x	y	\dot{x}	\dot{y}
A:	$\mathbf{X}_{LL_1-LL_1}^+$	Normalized	0.812255	0.0	0.0	0.248312
		SI (km, m/s)	312,230	0.0	0.0	254.418
B:	$\mathbf{X}_{LL_1-LL_1}^-$	Normalized	0.878585	0.0	0.0	−v0.281719
		SI (km, m/s)	337,728	0.0	0.0	−288.647
C:	$\mathbf{X}_{LL_1-DPO}^+$	Normalized	0.813049	0.0	0.0	0.247532
		SI (km, m/s)	312,536	0.0	0.0	253.618
D:	$\mathbf{X}_{LL_1-DPO}^-$	Normalized	0.890940	0.0	0.049050	−0.311179
		SI (km, m/s)	342,477	0.0	50.256	−318.830
E:	$\mathbf{X}_{DPO-DPO}^+$	Normalized	1.061692	0.0	0.0	0.403877
		SI (km, m/s)	408,115	0.0	0.0	413.809
F:	$\mathbf{X}_{DPO-DPO}^-$	Normalized	0.909845	0.0	0.0	−0.386264
		SI (km, m/s)	349,745	0.0	0.0	−395.762
G:	$\mathbf{X}_{DPO-LL_1}^+$	Normalized	1.056340	0.0	0.0	0.432104
		SI (km, m/s)	406,057	0.0	0.0	442.729
H:	$\mathbf{X}_{DPO-LL_1}^-$	Normalized	0.890940	0.0	−0.049050	−0.311179
		SI (km, m/s)	342,477	0.0	−50.256	−318.830

totically wraps off one orbit and onto the next as ϵ in Eq. (2.65) approaches 0; an infinite number of x -axis crossings precede the theoretical heteroclinic connection. The states D and H correspond to the x -axis crossings that are furthest from either host orbit. The states C and G correspond to the previous respective x -axis crossing. As one can see in Table 2-9, state C is approximately 306 km and 0.8 m/s away from state A, and state G is approximately 2058 km and 28.9 m/s away from state E. These state differences are small enough to proceed without difficulty.

The states summarized in Fig. 2-44 and Table 2-9 may be used to construct a sequence of states that represent any itinerary between the two given orbits. This sequence may then be converted into a series of patchpoints that may be inputted into a differential corrector in order to produce a continuous trajectory. For instance, the trajectory of a spacecraft in orbit about the LL_1 Lyapunov orbit may be represented by the sequence

$$\{\dots, A, B, A, B, \dots\}$$

A differential corrector may be used to convert this sequence into a continuous trajectory. If a mission designer wishes to transfer the spacecraft from the LL_1 orbit to the distant prograde orbit, the designer would construct the sequence

$$\{\dots, A, B, A, B, C, D, E, F, E, F, \dots\}$$

and input that sequence into the differential corrector. The differential correction process adjusts every state in the sequence to accommodate the slight differences between the states A and C to make the transfer continuous.

Table 2-10 provides two example sequences that may be used as inputs to a differential corrector in order to produce continuous trajectories with different itineraries. To demonstrate this process, the first sequence in Table 2-10 has been converted into patchpoints and processed by the multiple-shooting differential corrector described in Section 2.6.5.2. Table 2-11 displays the results of the differential correction process, comparing the states of the patchpoints before and after the process. One can see that the differential corrector adjusted each patchpoint away from the x -axis in order to produce a continuous trajectory, however, none of the patchpoints moved far. In this example, the differential corrector achieved a trajectory that met the requested continuity tolerances: the largest position and velocity discontinuities that were observed in any of the patchpoints along the final trajectory were less than 0.4 mm and 3.1×10^{-9} m/s, respectively.

2.6.12.2 Complex Periodic Orbits A complex periodic orbit may be constructed by repeating a given sequence of states *ad infinitum* and inputting that theoretical sequence into the differential corrector. For instance, the following sequence may be used to represent a periodic orbit that consists of two revolutions

Table 2-10 Two sequences that may be used as inputs to a differential corrector in order to produce continuous trajectories with different example itineraries. The letters correspond to the states summarized in Fig. 2-44 [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

Example 1		Example 2	
Sequence	Objective	Sequence	Objective
$\left. \begin{matrix} A \\ B \end{matrix} \right\}$	Traverse LL_1	$\left. \begin{matrix} A \\ B \end{matrix} \right\}$	Traverse LL_1
$\left. \begin{matrix} C \\ D \end{matrix} \right\}$	Transfer to DPO	$\left. \begin{matrix} C \\ D \end{matrix} \right\}$	Transfer to DPO
$\left. \begin{matrix} E \\ F \end{matrix} \right\}$	Traverse DPO (1)	$\left. \begin{matrix} G \\ H \end{matrix} \right\}$	Transfer to LL_1
$\left. \begin{matrix} E \\ F \end{matrix} \right\}$	Traverse DPO (2)	$\left. \begin{matrix} C \\ D \end{matrix} \right\}$	Transfer to DPO
$\left. \begin{matrix} G \\ H \end{matrix} \right\}$	Transfer to LL_1	$\left. \begin{matrix} E \\ F \end{matrix} \right\}$	Traverse DPO

Table 2-11 The results of a differential correction process that converted the first sequence presented in Table 2-10 into a continuous trajectory. The coordinates shown here are in the Earth–Moon synodic reference frame, relative to the Earth–Moon barycenter [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

Patchpoint	Initial State				Final State			
	x (km)	y (km)	\dot{x} (m/s)	\dot{y} (m/s)	x (km)	y (km)	\dot{x} (m/s)	\dot{y} (m/s)
A: $\mathbf{X}_{LL_1-LL_1}^+$	312,230	0.0	0.0	254.418	312,242	0.000	0.000	254.092
B: $\mathbf{X}_{LL_1-LL_1}^-$	337,728	0.0	0.0	-288.647	337,704	-0.003	0.034	-288.236
C: $\mathbf{X}_{LL_1-DPO}^+$	312,536	0.0	0.0	253.618	312,399	0.640	1.180	253.675
D: $\mathbf{X}_{LL_1-DPO}^-$	342,477	0.0	50.256	-318.830	342,553	12.782	51.463	-319.032
E: $\mathbf{X}_{DPO-DPO}^+$	408,115	0.0	0.0	413.809	407,079	19,004	6.862	427.761
F: $\mathbf{X}_{DPO-DPO}^-$	349,745	0.0	0.0	-395.762	349,595	-0.175	1.079	-393.771
E: $\mathbf{X}_{DPO-DPO}^+$	408,115	0.0	0.0	413.809	408,081	0.000	0.000	414.071
F: $\mathbf{X}_{DPO-DPO}^-$	349,745	0.0	0.0	-395.762	349,595	2.644	-1.094	-393.772
G: $\mathbf{X}_{DPO-LL_1}^+$	406,057	0.0	0.0	442.729	407,079	16,916	-7.171	427.755
H: $\mathbf{X}_{DPO-LL_1}^-$	342,477	0.0	-50.256	-318.830	342,557	-12.750	-51.431	-319.076
A: $\mathbf{X}_{LL_1-LL_1}^+$	312,230	0.0	0.0	254.418	312,555	0.000	0.000	254.418

about the LL_1 Lyapunov orbit, followed by one revolution about the distant prograde orbit, repeating itself indefinitely, where the orbits and letters are defined in Fig. 2-44

$$\{\dots, \underline{A, B, A, B}, \underline{C, D, G, H}, \underline{A, B, A, B}, \underline{C, D, G, H}, \dots\}$$

Figure 2-45 shows a plot of such a periodic orbit. One can see that a trajectory following a complex itinerary gets very close to one of its generating three-body orbits even with as few as two revolutions about the orbit.

Since each unstable three-body orbit exists in a family, where the characteristics of each orbit in the family vary continuously from one end of the family to the other, it is hypothesized that a complex periodic orbit also exists in a family. The family of any given periodic orbit is limited in extent to some range of parameters [46]. The extent of the family of complex orbits is also limited in extent, and it is hypothesized that the family may only extend through a range where each of its fundamental orbits and orbit transfers exists. Figure 2-46 shows several example complex periodic orbits that exist in the same family as the orbit shown in Fig. 2-45. Each of these orbits has a different Jacobi constant, but the same morphology.

2.6.12.3 Generalization The method demonstrated here has been illustrated by a very straightforward example, namely, the construction of orbit transfers between an LL_1 Lyapunov orbit, and a distant prograde orbit, two simple periodic three-body orbits. These orbits have been used because they are easily visualized and may be characterized using only a handful of states. Each state is placed at an x -axis crossing, although one can see in Table 2-11 that the states may be displaced during the differential correction process.

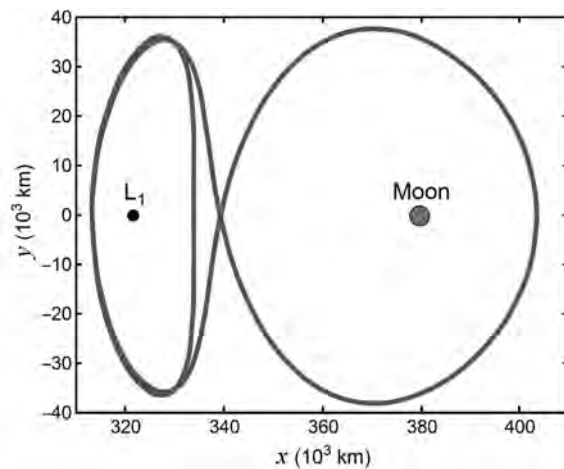


Figure 2-45 A complex periodic orbit that consists of two revolutions about the LL_1 Lyapunov orbit, followed by one revolution about the distant prograde orbit, repeating itself indefinitely. This orbit is viewed from above in the Earth–Moon synodic reference frame.

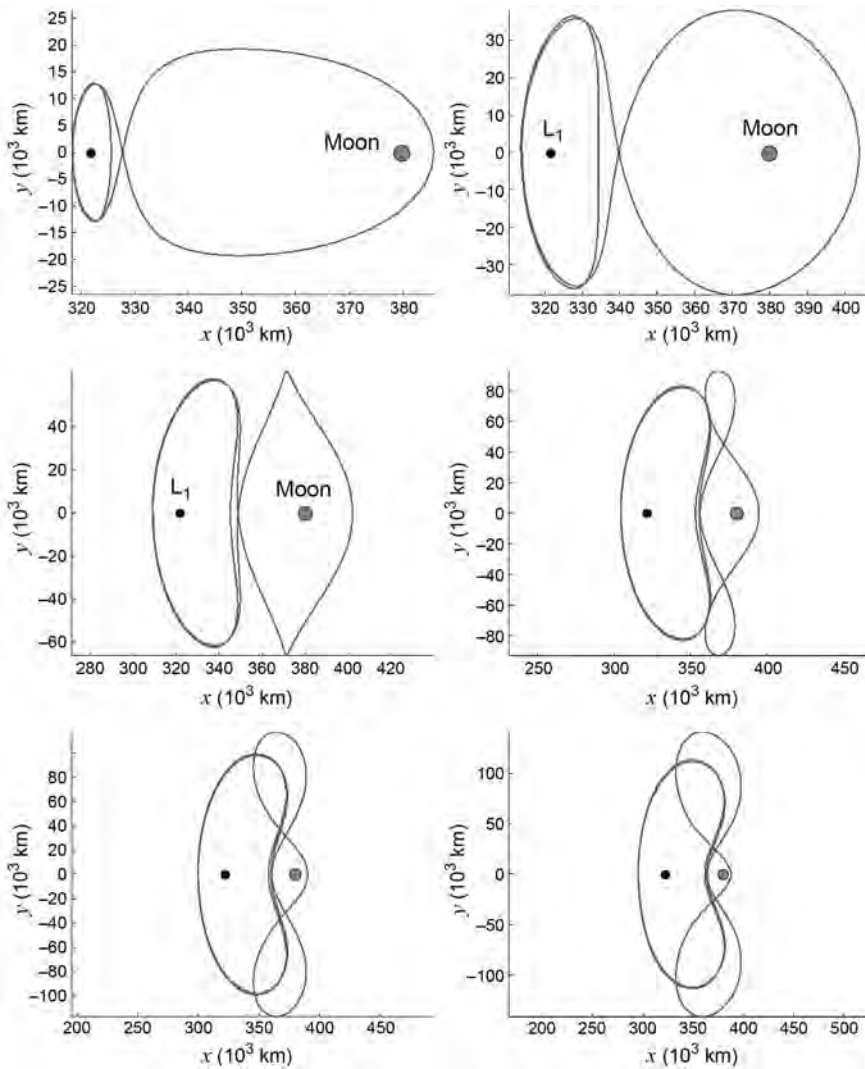


Figure 2-46 Several example complex periodic orbits that exist in the same family as the orbit shown in Fig. 2-45.

This method may certainly be applied to orbit transfers between other unstable three-body orbits, including non-symmetric orbits. In addition, a chain of orbits may certainly contain more than two different three-body orbits. Longer orbits and orbit transfers will likely require more states per segment for the differential corrector to converge. In that case, it is easier to visualize the problem by defining a sequence of states per segment and using symbols that represent sequences rather than individual

states. Table 2-12 provides an example where the states A–H given above have been mapped to four such sequences.

If one refers to Fig. 2-41, one notices that there are two low-energy transfers between the example Lyapunov orbits about LL_1 and LL_2 . One may construct a different sequence of states for each of those transfers, for example, $S_{LL_1-LL_2}^1$ and $S_{LL_1-LL_2}^2$, which may be constructed from three or more states, including an initial state and two intermediate states in order to keep the trajectory segment lengths short enough to permit the differential corrector to converge.

Figure 2-42 shows several low-energy transfers that exist from an orbit about LL_1 to an orbit about LL_2 that were generated using a Poincaré map. The transfer labeled (7) may be described as a complex chain that starts in an orbit about LL_1 , transfers to a DPO, remains in that orbit for a revolution, and then transfers from there to an orbit about LL_2 . This complex chain was identified using a Poincaré map, but it may be quickly generated by differentially correcting the series of states represented by the following sequence of

$$\{\dots, S_{LL1}, S_{LL1}, S_{LL1-DPO}, S_{DPO}, S_{DPO-LL2}, S_{LL2}, S_{LL2}, \dots\}$$

2.6.13 Discussion

This section introduced the tools that may be used to construct interplanetary transfers in the CRTBP using dynamical systems theory. It introduced the basic solutions of the CRTBP, including the Lagrange points and simple periodic orbits. It discussed several methods that may be used to build periodic orbits in the CRTBP. The stability of a trajectory or an orbit may be evaluated using the eigenvalues of the state transition or monodromy matrices. The state transition matrix is also very useful when implementing targeting tools such as the differential corrector. The unstable nature of many trajectories in the CRTBP leads to divergent behavior and chaos, but it also permits a mission designer to build low-energy transfers from one orbit to another. Mission designers trace structure in an orbit's stable and unstable manifolds and use that information to target a transfer to/from that orbit. Such transfers require

Table 2-12 The mapping of the states A–H to sequences [101] (*Acta Astronautica* by International Academy of Astronautics, reproduced with permission of Pergamon in the format reuse in a book/textbook via Copyright Clearance Center).

Sequence	States	Purpose
S_{LL1}	$\{A, B\}$	Traverse the LL_1 Orbit
$S_{LL1-DPO}$	$\{C, D\}$	Transfer from LL_1 to DPO
S_{DPO}	$\{E, F\}$	Traverse the DPO
$S_{DPO-LL1}$	$\{G, H\}$	Transfer from DPO to LL_1

very little energy and may be used to move a spacecraft a great distance around the three-body system without expending much fuel. These transfers are the basis for building ballistic transfers between the Earth and the Moon, which is the topic of the next few chapters.

2.7 TOOLS

Many tools are used in the design of low-energy lunar transfers. Dynamical systems methods and the corresponding tools, such as the differential corrector and Poincaré sections, are described earlier in this chapter. Other tools include numerical integrators and optimizers. These will be briefly described here.

2.7.1 Numerical Integrators

The two primary integrators used in the analyses contained in this work are the DIVA integrator [164–166] and a Runge-Kutta-Fehlberg seventh-order (RKF78) integrator with step-size control [167]. The DIVA integrator is currently implemented in both the Mission-analysis, Operations, and Navigation Toolkit Environment (MONTE) and libration point mission design tool (LTool) software (see Section 2.7.3) and has a rich heritage spanning more than three decades as an integrator for interplanetary missions at the Jet Propulsion Laboratory. It uses a variable-order Adams method for solving ordinary differential equations that has been written specifically for integrating trajectories. The RKF78 integrator is implemented in JPL's LTool. It allows for a variable step size as described by Fehlberg [167], and it is also widely used for astrodynamics and mission design.

2.7.2 Optimizers

Many mission designs presented in this book take advantage of the software package SNOPT (Sparse Nonlinear OPTimizer) [168, 169] to adjust the various parameters in order to identify solutions that require minimal amounts of fuel. SNOPT is written to use a particular sequential quadratic programming (SQP) method, one that takes advantage of the sparsity of the Jacobian matrix of the constraints of the system while maintaining a quasi-Newtonian approximation of the Hessian of the Lagrangian of the system. The details of the algorithms are beyond the scope of this discussion, except to say that they are written to be highly effective when applied to a system that has smooth nonlinear objective functions [169].

The objective functions and constraints studied here are indeed nearly smooth functions. There are two reasons why the functions studied in this paper are not perfectly smooth. First, the unstable nature of low-energy lunar transfers combined with finite-precision computers yields functions that have discontinuities. In general, these discontinuities are several orders of magnitude smaller than the trends being studied in this paper and are therefore ignored. Second, several objective functions studied in this paper involve iterative algorithms; there are discontinuities between a

set of parameters that require n iterations to generate a solution and a neighboring set of parameters that require $n + 1$ iterations to converge. The discontinuities observed are small relative to the topography in the state space; thus, SNOPT tends to work well in these studies.

2.7.3 Software

JPL's MONTE software [170] has been used for the majority of the analyses contained in this book. It provides an interface with JPL's DE421 Planetary and Lunar Ephemerides as well as integration using the DIVA propagator. JPL's LTool has been used for many of the computations involving libration orbits and their invariant manifolds. Targeting and optimization algorithms have been implemented in both sets of software for analyses in the CRTBP and in the ephemeris model. All of the coordinate frames described in Section 2.4 are accessed through the SPICE Toolkit in both software suites [171].

