The Genesis Trajectory and Heteroclinic Connections

Wang Sang Koon
Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125

Martin W. Lo
Jet Propulsion Laboratory
Navigation and Flight Mechanics Section
California Institute of Technology
Pasadena, CA 91109

Jerrold E. Marsden
Shane D. Ross
Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125

Abstract

Genesis will be NASA's first robotic sample return mission to return to the Earth. The purpose of this mission is to collect solar wind samples for two years in an L_1 halo orbit and return them to the Utah Test and Training Range (UTTR) for mid-air retrieval by helicopters. This requires the Genesis spacecraft to make an excursion into the region around L_2. This transfer between L_1 and L_2 requires no deterministic maneuvers and is provided by the existence of so-called heteroclinic cycles defined below. The Genesis trajectory was designed with the knowledge of the conjectured existence of these heteroclinic cycles. We now have provided the first semi-analytic construction of such cycles. The heteroclinic cycle provides several interesting potential applications for future missions. First, it provides a rapid low-energy dynamical channel between L_1 and L_2 such as used by the Genesis mission. Second, it provides a dynamical mechanism for the temporary capture of objects around a planet without propulsion. Third, a thorough understanding of this dynamics is essential for an optimal design of any constellations to study the magnetosphere region. Lastly, an understanding of the resonance structure of this dynamical regime may provide new strategies for low energy planetary captures at Mars or Europa.

1. Introduction

The key feature of the Genesis trajectory (see Fig. 1) is the Return Trajectory which makes a 3 million km excursion between L1 and L2 in order to reach UTTR (the Utah Test and Training Range) during daylight hours. The extraordinary thing about this 5 month excursion is that it requires no deterministic maneuvers! This transfer is called a heteroclinic connection between the L1 and L2 regions in dynamical systems theory.

Figure 1. The Genesis Trajectory in Sun-Earth Rotating Frame.

There is, in fact, a vast theory about heteroclinic dynamics which among other things is the generator of deterministic chaos in a dynamical system. Our recent work [Koon et al, 1999] provides a precise theory of how heteroclinic dynamics arise in the context of the planar circular RTBP (restricted three-body problem). In this paper, we apply this theory to explain how the Genesis Return Trajectory works. This provides the beginnings of a systematic approach to the design and generation of this type of trajectories. In the not too distant future, automation of this process will be possible based on this approach. The eventual goal is for the on board autonomous navigation of this type of low-energy Earth sample return missions. But in fact, as we will show, this dynamics affects a much greater class of new mission concepts.

To motivate the discussion and to provide an independent example from nature, we examine the orbit of the comet, Oterma, to see how this dynamics works in nature both to develop and to verify our theory. This is an important theme in our work, perhaps in science and engineering in general, to learn from nature because it seems that nature always has the best solution. Whatever we can glean from natural phenomena will contribute immeasurably to the development of new trajectory and mission concepts. In particular, the understanding of the structure of the heteroclinic and homoclinic orbits has given us new insights into the transport mechanisms within the Solar System and for space trajectory design.

Next, we explain the key theorem from Koon et al [1999] which we use to explain the "Temporary Capture Mechanism" in the astrodynamics context. It seems that the phase space region between L_1 and L_2 is full of dynamical channels like a complex system of tunnels or wormholes. These channels exists throughout the Solar System in a vast network connecting all the planets and their satellites (see Lo and Ross [1997, 1998]). Together, they provide the low energy transport which may be used for new mission concepts. We apply it to a new class of missions which we call "The Petit Grand Tour". The Petit Grand Tour combines the Temporary Capture Mechanism with the concept of the Interplanetary Network of Dynamical Channels to provide a low-energy mission to tour the moons of Jupiter (or Saturn) in a prescribed sequence by design.

2. Heteroclinic Connections and Cycles

The goal of the Genesis Mission is to return to UTTR all of the solar wind samples collected over two years in an L_1 halo orbit (see Lo et al [1998]). The mid-air retrieval of the ballistic SRC (Sample Return Capsule) by helicopters requires that the Entry must occur during daylight. But the natural dynamics of L_1 halo orbit requires a night-side return. In order to achieve the day-side Entry within a reasonable ΔV budget, an excursion into the L_2 region is necessary. This added about two months extra to the Return Phase.

A heteroclinic connection, $\mathcal{H}$, also called a heteroclinic trajectory (orbit), is an asymptotic trajectory which connects two periodic orbits which we denote by $\mathcal{A}$ and $\mathcal{B}$ for this discussion. In the event $\mathcal{A}$ and $\mathcal{B}$ are the same periodic orbit, $\mathcal{H}$ is called a homoclinic orbit. $\mathcal{H}$ is a theoretical construct of great importance both in theory and in practical applications. We examine some of the key features of these orbits. It takes $\mathcal{H}$ infinite time to wind off from $\mathcal{A}$ to transfer to $\mathcal{B}$. Once near the vicinity of $\mathcal{B}$, it takes $\mathcal{H}$ an infinite time to wind onto $\mathcal{B}$. They were studied intensely by Poincare (Barrow-Green [1991]) and was the key to his discovery of chaos in the 3 body problem. Practically, of course, we are never able to produce the real $\mathcal{H}$ just like we are never able to compute the real periodic orbits of any nonlinear system. However, what we are able to compute are neighboring trajectories, H's, which "shadow" $\mathcal{H}$ to any desired accuracy (within machine accuracy) nearly everywhere of importance to the particular problem at hand.

Now a slight diversion on our notation which we will keep to a minimum. We denote a heteroclinic orbit between $\mathcal{A}$ and $\mathcal{B}$ by $\mathcal{H}_{\mathcal{AB}}$ and a heteroclinic orbit between $\mathcal{B}$ and $\mathcal{A}$ by $\mathcal{H}_{\mathcal{BA}}$. In particular, a homoclinic orbit of $\mathcal{A}$ is denoted by $\mathcal{H}_{\mathcal{AA}}$. We distinguish the theoretical orbit, $\mathcal{H}$, and its numerical shadow, H, by the script and block fonts respectively.

Returning to our main discussion, when we have a heteroclinic orbit between $\mathcal{A}$ and $\mathcal{B}$ and a heteroclinic orbit between $\mathcal{B}$ and $\mathcal{A}$, the two orbits $\{ \mathcal{H}_{\mathcal{AB}} , \mathcal{H}_{\mathcal{BA}} \}$ is called a heteroclinic cycle. In particular, homoclinic orbits are already cycles. The importance of cycles both theoretically and practically will be discussed shortly below. And, they are very important indeed.

Of course, the existence of heteroclinic connections was generally known to the halo mission community. Typically when integrating an L_1 halo orbit for too long, it escapes the halo orbit and returns towards Earth and continues to wind around L_2. The WIND Mission was the first to use this heteroclinic behavior between L_1 and L_2 (Sharer et al, [1995]). Howell and Barden [1994] made a more formal study of heteroclinic connections and were able to find a free connection between a halo orbit and a lissajous orbit using numerical search.

Koon, Lo, Marsden, Ross [1999] studied the problem of PCRTBP (Planar Circular Restricted Three Body Problem) and used the more systematic and standard method of Poincare sections from dynamical systems theory to produce heteroclinic orbits between two Lyapunov orbits (periodic orbits around L_1 and L_2 in the plane). Although the method of Poincare section is also numerical in nature, it reduces the problem by 1 dimension thereby making the problem more tractable and bringing the problem one step closer to automation. Futhermore, there is a substantial theory and results on Poincare sections from dynamical systems theory which provide additional knowledge and insight into the specifics of the dynamics. This knowledge and insight provide the foundation for new mission concepts and for optimization of current mission concepts discussed in this paper.

We conclude this section by emphasizing the importance of these seemingly esoteric theoretical constructs, the $\mathcal{H}$ and even H orbits. Their importance in astrodynamics are two fold: (1.) Computation, (2.) New mission concepts. Their importance to computation is perhaps best illustrated by the process used to compute the Genesis halo orbit which we shall denote by A. We start the process with a theoretical model lissajous orbit, $\mathcal{A}$, specified by amplitudes and phase angles. We produce an analytic approximation, A_1 using a 3rd order analytic expansion. Next we produce a differentially corrected lissajous orbit, A_2, from A_1. Finally, starting with A_2, we apply the various mission constraints and differentially correct for the Genesis halo orbit, A. To summarize, the theoretical model orbit, $\mathcal{A}$, is the starting point from which practical orbits may be constructed via the continuation process using a series of numerical computations. In the same way, the Genesis Earth Return orbit was computed using heteroclinic orbits as initial models. Therefore, advances in the theory and computation of these orbits are essential to the simplification and eventual automation of this complex process of continuation. It is remarkable to think how Poincare was able to see all of these complex issues and actually perform continuation calculation of orbits without the benefit of modern computers. The discussion of their importance to new mission concepts is essentially the body of the rest of this paper.

2.1 The Three Body Problem

We start with the PCRTBP (Planar Circular Restricted Three Body Problem) as our first model of the mission design space, the equations of motion for which in rotating frame with normalized coordinates are:

$$x'' - 2y' = \Omega_x, \quad y'' + 2x' = \Omega_y, \qquad (1)$$

$$\Omega = \frac{x^2 + y^2}{2} + \frac{1-\mu}{r_S} + \frac{\mu}{r_J},$$

where subscripts denote partial differentiation in the variable and apostrophes after the variables are time derivatives. The variables r_S, r_J, are the distances from (x,y) to the Sun and Jupiter respectively.

The coordinates of the equations use standard PCRTBP conventions: the sum of the mass of the Sun and the Planet is normalized to 1 with the mass of the Planet set to μ; the distance between the Sun and the Planet is normalized to 1; and the angular velocity of the Planet around the Sun is normalized to 1. Hence in this model, the Planet is moving around the Sun in a circular orbit with period 2π. The rotating coordinates, following standard astrodynamic conventions, is defined as follows: the origin is set at the Sun-Planet barycenter; the x-axis is defined by the Sun-Planet line with the Planet on the positive x-axis; the xy-plane is the plane of the orbit of the Planet around the Sun.

Although the PCRTBP has 3 collinear libration points which are unstable in the cases of interest to mission design, we will examine only L_1 and L_2 in this paper. These equations are autonomous and can be put into Hamiltonian form with 2 degrees of freedom. It has an energy integral called the Jacobi constant which provides 3 dimensional energy surfaces:

$$C = -(x'^2 + y'^2) + 2\Omega(x,y) . \qquad (2)$$

The power of dynamical systems theory is that it is able to provide additional structures within the energy surface and characterize the different regimes of motions.

2.2 Examples from Nature: The Motion of Comets

The Jupiter family of comets exhibit many puzzling phenomena the most interesting of which to mission design is the *Temporary Capture Phenomenon*. Lo and Ross [1977] proposed another explanation based on observations of the stable and unstable manifolds of L_1 and L_2. Figure 2a. shows the stable manifolds as dashed curves, the unstable manifolds as solid curves for L_1 and L_2 of the Sun-Jupiter system in rotating frame. The Sun is labeled S, Jupiter is labeled J. In Figure 2b., the orbit of the comet Oterma is overlayed in color. The orange segment is in the Exterior Region, outside of Jupiter's orbit; the blue segment is in the Interior Region, inside of Jupiter's orbit. Notice how well Oterma's orbit fits with that of the manifolds of L_1 and L_2. Lo and Ross argued that this suggest that the comet orbit is under the control of the invariant manifold structure of the Lagrange points. The term invariant manifold structure is a catch-all phrase for the entire structure of periodic and quasiperiodic orbits around the Lagrange points and all of their associated invariant manifolds such as the stable and unstable manifolds of the periodic orbits. Recall a manifold is simply a mathematical term for higher dimensional surfaces. An invariant manifold in dynamical systems theory is a special manifold consisting of orbits; hence a point on the invariant manifold will forever remain on the manifold under the flow of the equations of motion. The Lagrange points are examples of 0-dimenisonal invariant manifolds. A periodic orbit is an exmaple of a 1-dimensional invariant manifold. The stable manifold of a Lyapunov orbit is an example of a 2-dimensional manifold. Its energy surface in the PCRTBP is an example of a 3-dimensional invariant manifold.

Figure 2a. The Stable and Unstable Manifold of Jupiter's L_1 and L_2.
2b. The Orbit of Oterma Overlaying the Manifolds of Jupiter's L_1 and L_2.

In Figure 3a. the orbit of comet Gehrels3 is overlaid against the manifolds of Jupiter's L_1 and L_2. In Figure 3b., a close-up in the Jupiter Region shows how Gehrels3 goes into a halo orbit for one revolution around L_2 before capturing into Jupiter orbit for several revolutions. Once again, the manifolds match closely with the comet orbit. Furthermore, the Temporary Capture of the comet by Jupiter suggests the possibilities of low-energy capture for interplanetary missions.

Figure 3a. The Orbit of Gehrels3 Overlaying the Manifolds of Jupiter's L_1 and L_2.
3b. The Orbit of Gehrels 3 in the Jupiter Region Showing Temporary Captures and Halo Orbits.

Based on the invariant manifold approach suggested by Lo and Ross [1997], Koon et al [1999] provides a systematic and mathematically rigorous explanation of this dynamics. In addition to a more complete global qualitative picture of the dynamics, it has computational and predictive capabilities. It provides an algorithm based on standard Poincare Section methods for the computation of heteroclinic orbits. It provides the rudiments for the calculation of transport coefficients based on heteroclinic lobe dynamics theory. Previous mission design work using heteroclinic dynamics were based on ad hoc numerical search and exploration. But the new computation tools of Koon et al [1999] enables the mission designer to construct heteroclinic trajectories in a systematic fashion instead of a blind search. Much research and development work remains before this process may be fully automated.

2.3 Orbit Classes Near L_1 and L_2

We now review the work of Conley [1968], Moser [1973], and McGehee[1969] which provides an essential characterization of the orbital structure near L_1 and L_2. McGehee also proved the existence of homoclinic orbits in the Interior Region. Llibre, Martinez, and Simo [1985] computed homoclinic orbits of L_1 in the Interior Region. They further extended McGehee's results and proved a theorem using symbolic dynamics for orbital motions in the Interior Region. The key result in Koon et al [1999] is the completion of this picture with the computation of heteroclinic cycles in the Jupiter Region between L_1 and L_2. We will refer to the various regions by the following short hand: S for the Interior Region which contains the Sun, J for the Jupiter Region which contains the planet, X for the Exterior Region.

Figure 4 below schematically summarizes the key results of Conley, Moser, and McGehee. For energy value just about that of L_2, the Hill Region is projection of the energy surface from the phase space onto the configuration space, i.e. the xy-plane. This is represented by the white space in Figure 4a. The grey region is the energetically forbidden. In other words, with the given energy, our spacecraft can only explore the white region. More energy is required to enter the grey Forbidden Region.

Figure 4a. The Hills Region Connecting the Interior Region, the Capture Region, and the Exterior Region.
4b. Expanded View of the L_2 Region with 4 Major Classes of Orbits:
Black: Periodic Orbits, Green: Asymptotic Orbits
Red: Transit Orbits, Blue: Non-Transit Orbits

Figure 4b blows up the L_2 Region to indicate the exitence four different classes of orbits. The first class is a single periodic orbit with the given energy, the planar Lyapunov orbit around L_2. The second class represented by a Green squiggle is an asymptotic orbit winding onto the periodic orbit. This is an orbit on the stable manifold of the Lyapunov orbit. Similarly, although not shown, are orbits which wind off the Lyapunov orbit to form the unstable manifold. The third class, represented by red orbits, are transit orbits which pass through the Jupiter Region between the S and X Regions. Lastly, the fourth class in blue consists of orbits which are trapped in the S and X Regions.

Let us examine the stable and unstable manifold of a Lyapunov orbit as shown in Figure 5 below. Of course, only a very small portion of the manifolds are plotted. Notice the X-pattern formed by the manifolds, reminiscent of the X-pattern of the manifolds of the Lagrange points. It is precisely in this sense that we say the manifolds of the Lagange points are "genetic", in that they characterize the shapes and the dynamics of things to come when more complexity such as periodic orbits and their manifolds are introduced. Thus we study the simple 1-dimensional manifolds of L_1 and L_2 to gain some understanding of the nature of the complex dynamics of the full invariant manifold structure of the region.

Figure 5. The Stable and Unstable Manifold of a Lyapunov Orbit.

Notice that the 2-dimensional tubes of the manifolds of the Lyapunov orbits are separatrices in the 3-dimensional energy surface! By this we mean the tubes separate different regions of motion within the energy surface. Referring back to the schematic diagram, Figure 4b, we notice that the Red Transit Orbits pass through the oval of the Lyapunov orbit. This is no accident, but an essential feature of the dynamics on the energy surface. Lo and Ross [1997] had referred to L_1 and L_2 as gate keepers on the trajectories since the Jupiter comets must transit between the X and S regions through the J region and always seem to pass by L_1 and L_2. Chodas and Yeomans [1996] noticed that the comet Shoemaker-Levy9 passed by L_2 before it's crash into Jupiter. These tubes are the only means of transit between the different regions in the energy surface! In fact, all this was already known to Conley and McGehee.

2.4 The Homoclinic-Heteroclinic Chain

By putting all of these results together, we are able construct a complete chain as show in Figure 6: start with a homoclinic cycle (Blue) in the Interior Region, go to a heteroclinic cycle in the Capture Region (Magenta), and finally end with a homoclinic cycle (Orange) in the Exterior Region. The pair of Lyapunov orbits around L_1 and L_2 which generated this chain are in black. The existence of this chain has many important implications for mathematics, astronomy, and astrodynamics. Let us take a moment to see heuristically exactly what this chain means. We have essentially produced a series of asymptotic trajectories that connect the S, J, X regions. So what? Since these theoretical orbits take infinite time to complete their cycles, of what use are they?

Figure 6a. Jupiter's Homoclinic-Heteroclinic Chain.
6b. The Lyapunov Orbits (Black) and the Heteroclinic Cycle (Magenta).

Well, recall that large body of theory and results in dynamical systems theory relating to heteroclinic orbits mentioned earlier? Here is where we cash in our chips after hitting the jack pot. It turns out one of the sources of deterministic chaos in a dynamical system is precisely the existence of homoclinic and heteroclinic cycles. This was known to Poincare and gave him enormous difficulties. Basically, when these cycles exist, it implies that the stable and unstable manifolds have infinite number of intersections creating what is known as the homoclinic/heteroclinic tangle. This is truly a mess. The existence of this tangle means that very random transitions between the S, J, X set of regions can occur using the chain as the template for the transition. In other words, a comet could orbit the Sun in the X Region for many years, then suddenly changes its orbit to the S Region. Of course, to do this, it must transit through the J Region where it might get caught by Jupiter for a couple of orbits. It might also be caught by L_1 or L_2 doing a few revolutions of a halo orbit. Then it leaves via the L_1 Region to enter the S Region and orbit the Sun. This, of course, is exactly the itinerary of Gehrels3, Oterma, and a host of other comets. This dynamics is completely explained by the tangle associated to this chain.

But actually an even more precise result is proved in Koon et al [1999] using symbolic dynamics, a simple technique in dynamical systems theory. The basic idea is as follows. We want to characterize the dynamics by following it in space. But, the detail trajectory is too complicated. Suppose we divide the space into 3 regions such as S, J, X in Figure 7. Let's just track when the trajectory is in each of the 3 regions. Thus a trajectory is characterized by an infinite sequence (..., X, J, S, J, X, ...) indicating the "itinerary" of the trajectory. Certain sequences such as (..., X, S, ...) are impossible because as we know to go from X to S, the trajectory must pass through J. We call the set of all possible trajectories, Admissible Trajectories. The main theorem in Koon et al [1999] states that given any admissible itinerary, (..., X, J, S, J, X, ...), there exists a natural orbit whose whereabouts matches this itinerary. Here naturality implies no ΔV is required, a free energy transfer all the way! In fact, we can even specify the number of revolutions the trajectory makes around the Sun, Jupiter, L_1 or L_2! And this for an infinite sequence going back and forth between the S and X Regions!

Figure 7. The Symbolic Dynamics of Transitions Between the S, J, X Regions.

2.5 The Numerical Construction of Orbits with Prescribed Itinerary

At this point, skeptics will no doubt recall that mathematical existence proofs are worth very little for real engineering problems. This observation is quite mistaken. We use the Genesis trajectory design as an example. When we first studied the Genesis problem, what guided us in our mission design was the knowledge that there is heteroclinic behavior between the L_1 or L_2 regions. Thus the knowledge of the conjectured existence of these cycles provided the necessary insight for us to search in the design space to find the desired solution. Furthermore, our knowledge of heteroclinic orbit theory, though much less complete than it is today, provided the basic algorithms for the numerical search which produced the Genesis trajectory. We knew you had to compute periodic orbits at L_1 or L_2 and produce a transfer between them as a first step to find the Return Trajectory for Genesis. Once a heteroclinic-like orbit is constructed, perhaps studded with ΔV's, this orbit provided the starting point for our differential correction process which continued the orbit to eventually produce the 6 m/s ΔV mission!

Hence existence proofs and theory do provide invaluable, necessary insight to solve very practical engineering problem even when the computational machinery associated with the theory has not been developed.

A second point is the fact that our theory actually provides a completely systematic method for the numerical construction of orbits with arbitrary prescribed itinerary! Figure 8 below provides some details on how the heteroclinic orbits may be found.

Suppose we wish to compute a heteroclinic orbit from L_2 to L_1. We start with two periodic orbits of the same energy around L_1 or L_2 depicted in black in Figure 2a. We compute the unstable manifold (red) of the L_2 periodic orbit; we compute the stable manifold (green) of the L_1 periodic orbit. We find the intersection between the two manifolds at a convenient location such as the solid black line through Jupiter in Figure 2a. The solid black line actually represents a plane in phase space, say the {y, y'} - plane. When we intersect the stable manifold with this plane, we expect to get a distored green circle; similarly, the unstable manifold will intersect this plane in a distorted red circle. This is exactly what is shown in Figure 8b. This is called the Poincare Section, or the Poincare Cut. It has reduced our manifold (surface) intersection problem by 1 dimension into a curve intersection problem. This is a much simpler problem. We see that there are two intersections. Taking one of these, integrating this state backwards and forwards towards the periodic orbits around L_1 and L_2 produces the heteroclinic orbit in Figure 8c.

Figure 8a. The Intersection of Stable and Unstable Manifolds in the J Region.
8b. The Poincare Section of the Manifolds and Their Intersections.
8c. The Heteroclinic Orbit Generated from the Intersection.

Notice, what we have constructed above is the symbolic sequence (X; J, S). The semi-colon divides the past from the present. We came from the X Region; we are currently at the J Region and we will transfer to the S Region. In Figure 9, we show the (X; J, S) sequence graphically. Figure 9a. depicts the manifold tubes as in Figure 8a. Figure 9b. magnifies the intersection of the manifolds in the Poincare map. Recall that the invariant manifold tubes separate the transit orbits from the non-transit orbits. In other words, as we state earlier, all orbits entering the J Region from X at this energy level, must enter through the unstable manifold tube of the Lyapunov orbit! Hence, the red circular curve in the Poincare Section shows all orbits of the sequence (X; J). Similarly, the green circular curve captures all orbits leaving the J region to enter the S region at this energy level, the (J; S) sequence. Their intersection is exactly the (X; J, S) sequence, highlighted in yellow. And since hamiltonian systems preserve area, by comparing the area of these curves and intersections in the appropriate coordinates, we can actually compute the transition probability from one region of phase space to another! For a little more complexity, in Figure 10 we show an orbit with the itinerary (X, J; S, J, X).

Figure 9a. The Intersection of the Stable and Unstable Manifolds of Periodic Orbits.
9b. The Poincare Section of the Manifold Intersection. The Yellow Region Depicts the (X; J, S) Sequence.
Figure 10a. The (X, J; S, J, X) Orbit.
10b. The Details of the (X, J; S, J, X) Orbit in the J Region.

These observations merely scratch the surface of the transition probability calculus which is possible using this technique. Thus far from an esoteric mathematical curiosity, symbolic dynamics can be a very useful computational tool when viewed in this context. This remarkable theory is known as "lobe dynamics" in dynamical systems theory. It was developed about 10 years ago and is currently an active area of research.

3. Applications to the Genesis Mission

In Figure 11 we computed the chain for two Lyapunov orbits with the Jacobi energy of the Genesis halo orbit. Although the resulting heteroclinic orbit in the blow-up Figure 11b has an extra loop around the Earth, the general characteristics of the Return Orbit shadows the heteroclinic orbit in the gross details. As indicated in Bell, Lo, & Wilson [1999], the influence of the Moon is crucial to the Genesis trajectory. Perhaps the role of the Moon is to pull up the Magenta orbit closer to the Moon's orbit to produce the actual Earth Return Orbit. Also, since the Genesis orbit is fully 3-dimensional, our 2-dimensional theory may be missing important elements of the dynamics.

Figure 11a. The Homoclinic/Heteroclinic Chain for the Genesis Orbit.
11b. Detail Blow-Up of the Earth Region of the Genesis Chain.

One of the findings of Bell et al [1999] is the surprising robustness of the Genesis trajectory. Even when critical maneuvers were missed, the trajectoy is still able to submit to correction and return to UTTR within the 450 m/s total ΔV budget. In fact, given the now understood chaotic nature of this orbit, how is it possible to compute any orbits at all starting from guesses using pieces of the invariant manifolds? Why should the orbits seemingly cling to the surface of the invariant manifolds? What keeps the orbit shadowing the manifolds despite its sensitivity? This is explained heuristically by the lobe dynamics in a phenomenon technically called "stickiness". The invariant manifolds are sticky; this means orbits close to the invariant manifolds tend to remain there for a long period of time. This is because in the infinite homoclinic/heteroclinic tangle, when two manifolds intersect, their intersection creates pockets called lobes. Orbits in these lobes are trapped and remain in the lobe structure which is close to the invariant manifolds even after a long time. Hence it is the lobe dynamics which creates this stickiness; the stickiness in turn, produces the unexpected stability seen in the Genesis orbit as well as in ordinary comet orbits. It explains why the comets change orbits only intermittently. Once it gets near a resonance, it tends to stay there for sometime. Then inexplicably, it transitions to another resonance, perhaps changing the X, J, or S regions. This intermittency is explained by the lobe dynamics. The transition calculus provided by the lobe dynamics will permit us someday to compute the probabilities to characterize the intermittent behavior.

The fact that chaotic orbits like to stick around the resonances is illustrated by Figure 12, the mean motion resonance diagram created from a Poincare Section of Jupiter's L_1 manifolds in the S Region. Each dot in the diagram represents one orbit of the L_1 manifold around the Sun. An animated version would show how the dots tend to circulate around the resonances for long periods of time, indicating stickiness. Then jumping to another resonance. Eventually, the swiss cheese plot is generated. The semimajor axis has jumped between 2 to 5 AU over the time interval of this plot which is about 1 million years. The relation between resonance and invariant manifolds is a deep one, not entirely understood. For instance, it is well known that Jupiter comets tend to transition from X to S Region using the 2:3 resonance in the X Region, to the 3:2 resonance in the S Region. An (n:m) resonance means the comet orbits the Sun n times while Jupiter orbits the Sun m times during approximately the same period of time. As Figure 12 shows, the 3:2 resonance is indeed prominent in the Jupiter resonance diagram. More details regarding the resonance transition and heteroclinic orbits is given in Koon et al [1999].

Figure 12. The Resonance Characteristics of the Jupiter L_1 Manifold.

While the nominal trajectory for Genesis seems robust and malleable, finding the initial orbit was extremely difficult and time consuming. This suggests that given a sufficiently severe change in the orbit due to contingency problems, finding a new Return Trajectory will be very difficult. Our experience working on this trajectory design for the past three years indicates that when this trajectory goes wrong, it is very hard to fix. Unlike conventional halo orbit missions where the specific halo orbit is of no concern, so long as the spacecraft remains in the general vicinity of the Lagrange point, the mission objectives can be achieved. Genesis require the return of the solar wind samples precisely to UTTR in daylight. The combination of the UTTR target with daylight entry severely constrains the design problem. For example, the role of the moon in the Genesis orbit design was completely unplanned (see Bell et al [1999]). The Moon, in fact, was purposely avoided to eliminating the difficulty of phasing with the Moon. In the end, it could not be avoided.

A deeper understanding of the dynamics behind the Genesis Return Trajectory could greatly alleviate this problem. Clearly, Bell et al [1999] and this paper shows that a thorough investigation of the theory of heteroclinic orbits with lunar perturbations in the Sun-Earth system is critical. But of even greater importance is to have the proper tools at hand which are responsive to the demands of the many potential contingency situations that may arise. The development of the proposed JPL's LTool (Libration Point Mission Design Tool) is a respose to these challenges. With a deeper understanding of the fundamental dynamics, automation of the design process may be possible since the invariant manifold structures and the various computation algorithm associated with them are well defined, at least theoretically.

It is a general rule of thumb for these highly nonlinear trajectory missions, and perhaps for all missions, that the role of contingency planning is critical to the success of the mission. Invariably, something always goes wrong in a mission. And what goes wrong is never what you expect. In order to prepare for these challenges, the best bet is to study the orbit design space in order to find out what options are available. Then having a good, flexible design tool to be able to quickly implement these options will greatly enhance the mission and reduce the risk of failure.

4. Applications to Other Missions

Many potential applications of the dynamics of the chain are possible. The Genesis Mission is a prime example of an application of the heteroclinic connection between L_1 and L_2. The homoclinic orbits are very similar to the SIRTF-type heliocentric orbits. Clearly missions in the Earth Region between L_1 and L_2, those going to the Moon, or the extended Magnetosphere can all benefit from using this dynamics. We leave these mission applications to future papers. Instead, we want to introduce the "Petit Grand Tour" concept to complete this paper.

The Petit Grand Tour is a tour of the Jovian satellites. But unlike previous flyby tours, the concept here is to linger at each satellite in a temporary captured orbit for a prescribed number of orbits before moving on to the next satellite. The temporary capture orbits can be constructed using the heteroclinic cycles described above. But the intersatellite transfers requires a different mechanism. Figure 13 plots the L_1 and L_2 manifolds of the Galilean satellites, showing intersections between the L_1 manifold of the outer satellite with the L_2 manifold of the inner satellite. Note, this intersection is in configuration space only. A ΔV may be necessary to effect the transfer from the L_1 manifold of one satellite to the L_2 manifold of the next satellite. This network of dynamical channels was discovered by Lo and Ross [1997]. Using this dynamical network to leap from satellite to satellite, and using the heteroclinic cycles to effect low-energy temporary captures at a satellite, the Petit Grand Tour concept is thereby complete.

Figure 13. The Network of Interconnecting Dynamical Channels Generate by the Invariant Manifolds L_1 and L_2 of the Galilean Satellites of Jupiter.

In Figure 14 we illustrate a segment of the Petit Grand Tour of Jovian moons. We prescribe 1 orbit around Ganymede, leave Ganymede via the L_1 unstable manifolds, transfer to the Europa L_2 stable manifolds using a ΔV, then almost get into a Lyapunov orbit around Europa's L_2, and finally capture into Europa orbit for 4 orbits. The trajectory is integrated using the planar restricted bicircular problem (PRBCP). In this model, both Europa and Ganymede orbit Jupiter in circular orbits with no gravitational effects on one another. The ΔV savings is a little more than half that required for a Hohmann transfer between Ganymede and Europa, although the trajectory was not optimized in any way. This preliminary design is illustrative of the types of mission which as possible using these techniques.

Figure 14a. One Orbit Around Ganymede in Rotating Coordinates.
14b. The Transfer from Ganymede to Europa Via the Invariant Manifold Intersections..
14c. Temporary Capture by Europa into 4 Orbits.

5. Future Work

The work presented in this paper represents an initial foray into chaotic dynamics of the homoclinic/heteroclinic chain in the 3 body problem. There are many directions where this work may be continued. We mention a few of the most important problems and applications in astrodynamics.

Extension of the 2D heteroclinic cycle to 3D is the most important problem for astrodynamic applications since halo orbits and not Lyapunov orbits are the ones of most interest to missions. The second problem is the systematic study of Earth-Return/Collision orbits. The Genesis orbit, for example, is an Earth collision orbit not all that different from the spectacular Shoemaker-Levy9 orbit. The third problem is the interactions with the Moon. Here we speak of the interactions of the Sun-Earth Lagrange point dynamics with that of the Earth-Moon Lagrange point dynamics. This interaction provided the low energy capture which rescued the Hitten mission. Similarly, the recent Hughes satellite rescue mission also used this dynamics. The fourth problem, a more technical problem, is to combine dynamical systems theory with optimal control methods. It is hoped that the many difficulties which fact optimal control problems may be alleviated if the dynamics of the problems were taken more into consideration. For example, we are currently working on targeting the stable manifold to compute an optimal transfer into a halo orbit. Lastly, perhaps the most important problem in this field currently, is the development of good software tools to perform the analysis.

Acknowledgements:

We thank Gerard Gomez and Josep Masdemont for many helpful discussions and for sharing their wonderful software tools with us. We thank Paul Chodas for discussions on the SL9 and NEO orbits. We also thank the following colleagues for helpful discussions and comments: Brian Barden, Julia Bell, Peter Goldreich, Kathleen Howell, Angel Jorba, Andrew Lange, Jaume Llibre, Martinez, Richard McGehee, William McLaughlin, Linda Petzold, Nicole Rappaport, Ralph Roncoli, Carles Simo, Scott Tremaine, Stephen Wiggins, and Roby Wilson.

This work was carried out at the Jet Propulsion Laboratory and California Institute of Technology under a contract with National Aeronautics and Space Administration. In addtion, the work was partially supported by the Caltech President's fund, the NASA Advanced Concepts Research Program, the Genesis Mission, and NSF grant KDI/ATM-9873133.

References:

Barrow-Green J., [1997] Poincare and the Three Body Problem, American Mathematical Society

Belbruno, E., B. Marsden [1997] Resonance Hopping in Comets, The Astronomical Journal 113(4),1433-1444

Bell, J., M. Lo, R. Wilson, [1999] AAS/AIAA Astrodynamics Specialist Conference, Girdwood Alaska, August 16-20, 1999

Chodas, P., D. Yeomans, [1996] The Orbital Motion and Impact Circumstances of Comet Shoemaker-Levy 9, in The Collision of Comet Shoemaker-Levy 9 and Jupiter, Edited by K.S. Noll, P.D. Feldman, and H.A. Weaver, Cambridge Unviersity Press, 1996

Conley, C. [1968] Low Energy Transit Orbits in the Restricted Three-Body Problem, SIAM J. Appl. Math. 16, 732-746

Howell, K., D. Mains, B. Barden, [1994] Transfer Trajectories from Earth Parking Orbits to Sun-Earth Halo Orbits, AAS/AIAA Spaceflight Mechanics Meeting, Cocoa Beach, FL, Feb. 14-16, 1994

Koon, W.S., M. Lo, S. Ross, J. Marsden, [1999], Heteroclinic Connections between Liapunov Orbits and Resonance Transitions in Celestial Mechanics, draft

Llbre, J., R. Martinez, C. Simo, [1985] Transversality of the Invariant Manifolds Associated to the Lyapunov Family of Periodic Orbits Near L2 in the Restricted Three-Body Problem, Journal of Differential Equations 58, 1985, 104-156

Lo, M., et al, [1998] Genesis Mission Design, AAS/AIAA Astrodynamics Conference, Boston, MA, August 1998, Paper AIAA 98-4468

Lo, M., and S. Ross, [1997], SURFing the Solar System: Invariant Manifolds and the Dynamics of the Solar System, JPL IOM 312/97

Lo, M., and S. Ross, [1998], Low Energy Interplanetary Transfers Using Invariant Mnaifolds of L_1 and L_2 and Halo Orbits, AAS/AIAA Space Flight Mechanics Meeting, Monterey, CA, Feb 9-11, 1998

McGehee, R., [1969] Some Homoclinic Orbits for the Restricted Three Body Problem, Ph.D. Thesis, University of Wisconsin

Moser, J. [1973] Stable and Random Motions in Dynamical Systems with Special Emphasis on Celestial Mechanics, Princeton University Press

Sharer, P., H. Franz, D. Folta, [1995] WIND Trajectory Design and Control, Spaceflight Dynamics (Colloque 95), CEPAD, Toulouse, 1995, pp.859-870

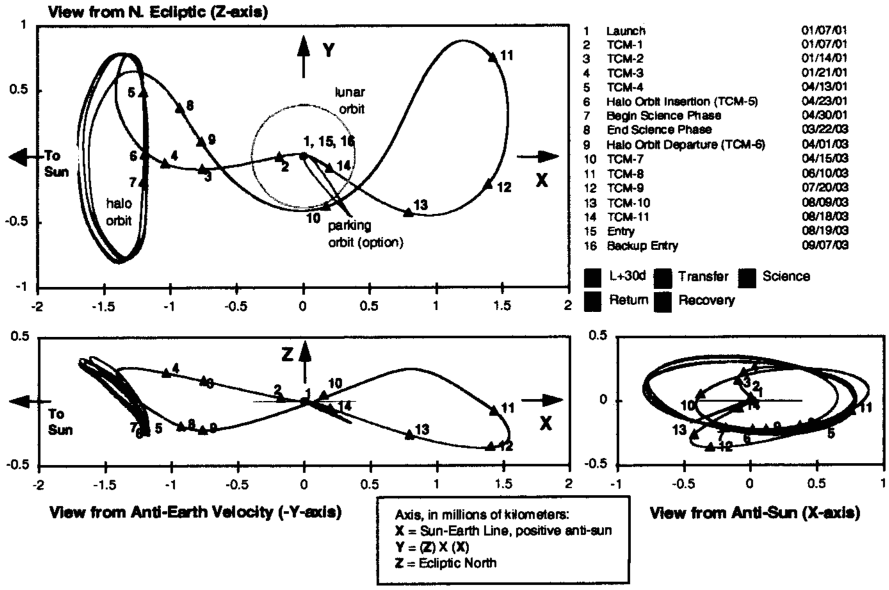


Figure 1. The Genesis Trajectory in Sun-Earth Rotating Frame.

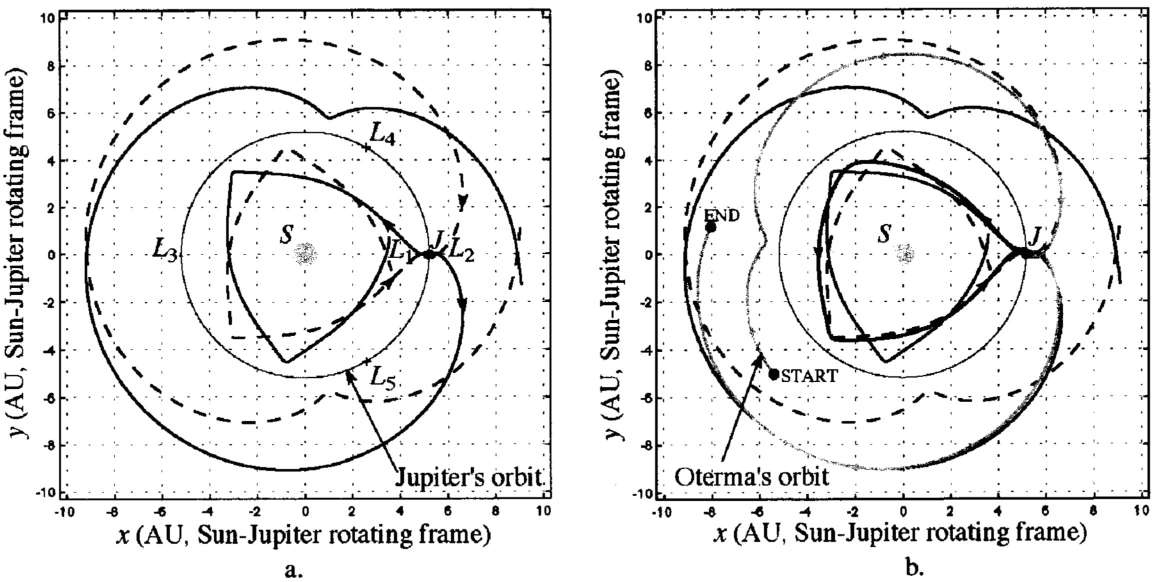
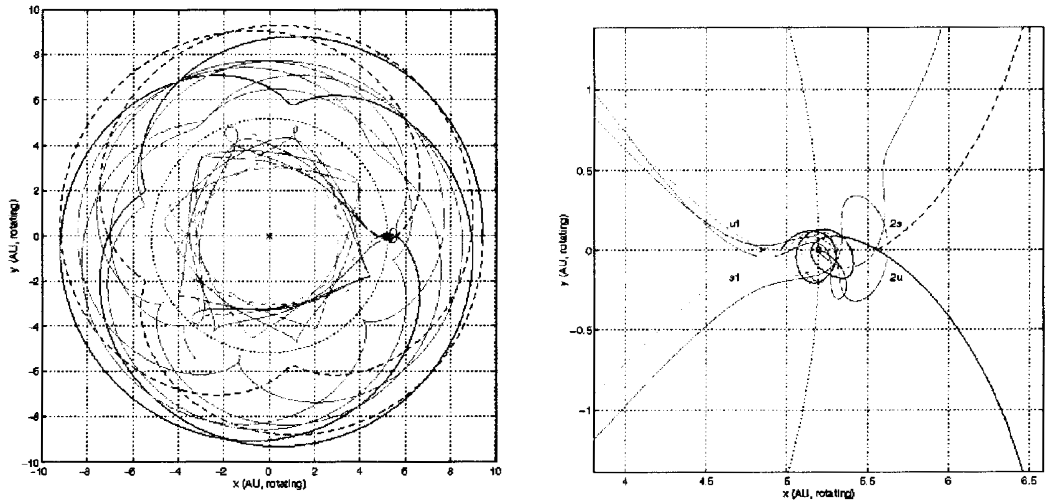
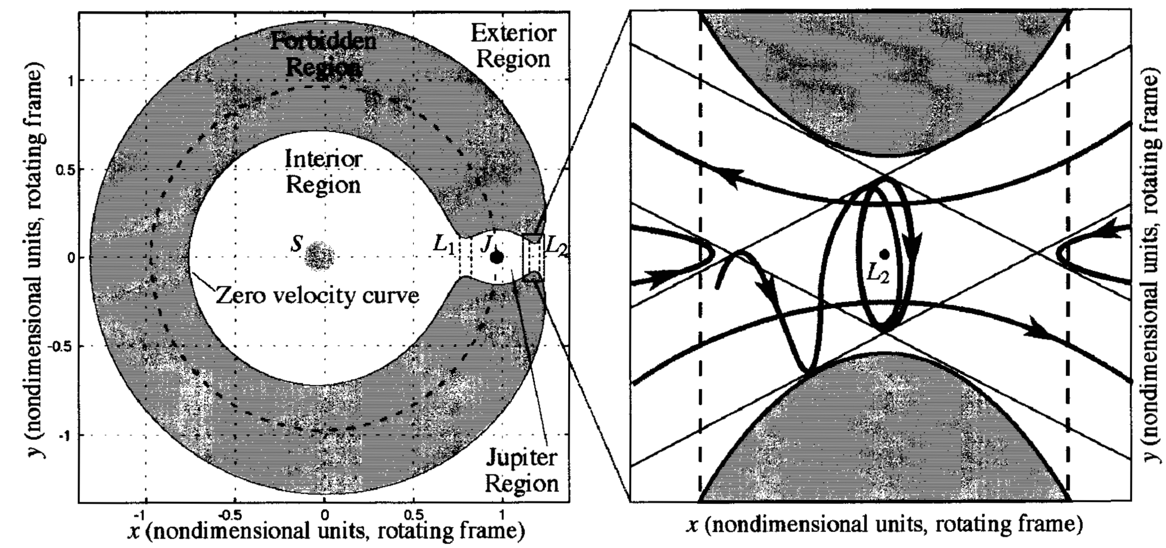


Figure 2a. The Stable (dashed curve) and Unstable (solid curve) Manifold of Jupiter's L_1 and L_2.
2b. The Orbit of Oterma Overlaying the Manifolds of Jupiter's L_1 and L_2.



a. b.

Figure 3a. The Orbit of Gehrels3 Overlaying the Manfiolds of Jupiter's L_1 and L_2.
3b. The Orbit of Gehrels3 in Jupiter Region Showing Temporary Captures and a Halo Orbit.



a. b.

Figure 4a. The Hill's Region Connecting the Interior (S), Jupiter (J), and Exterior (X) Regions.
4b. The Orbit of Gehrels3 in Jupiter Region Showing Temporary Captures and a Halo Orbit.

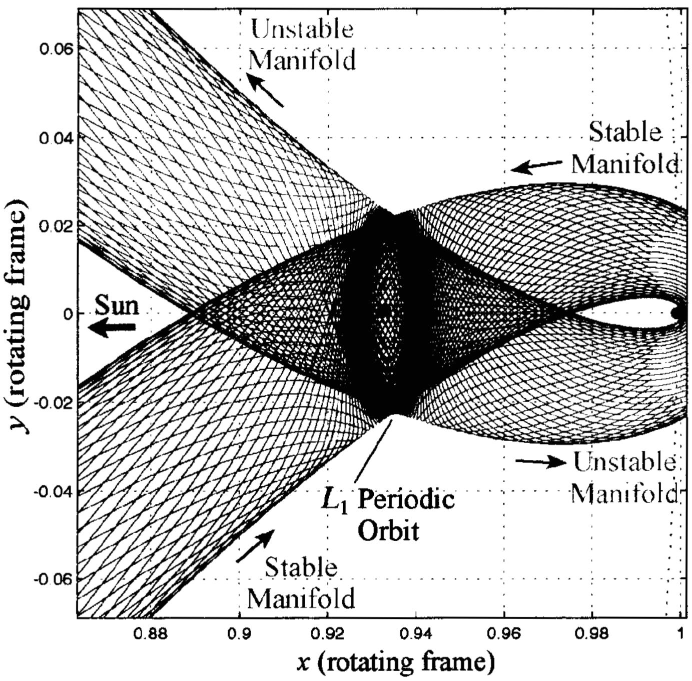


Figure 5. The Stable (Green) and Unstable (Red) Manifolds of a Lyapunov Orbit.

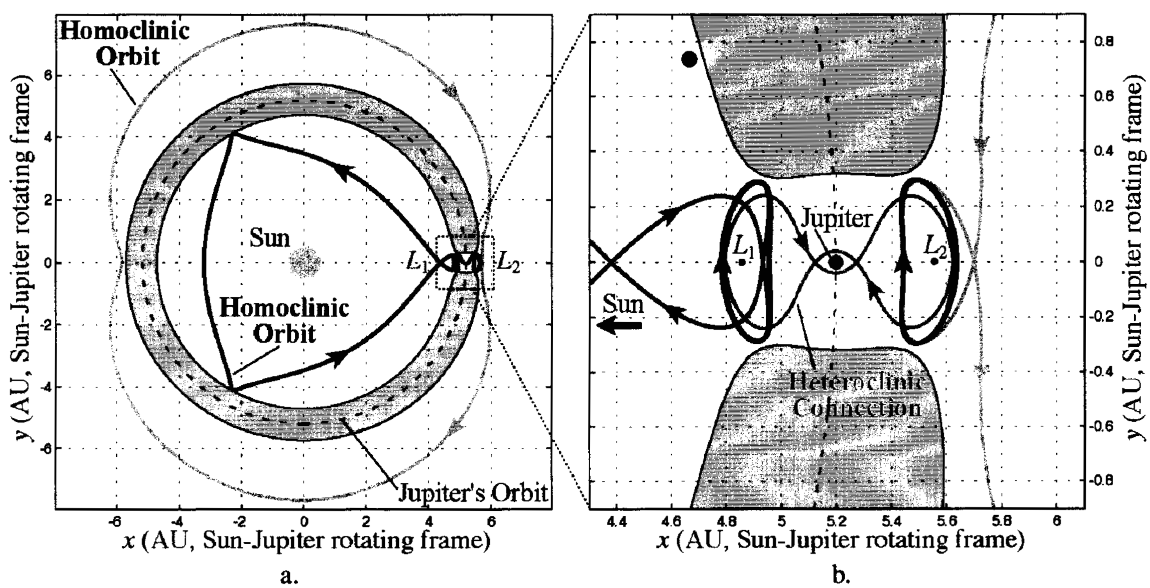


Figure 6a. Jupiter's Homoclinic (Orange, Blue) and Heteroclinic (Magenta) Chain.
6b. The Lyapunov Orbits (Black) and the Heteroclinic Cycle (Magenta).

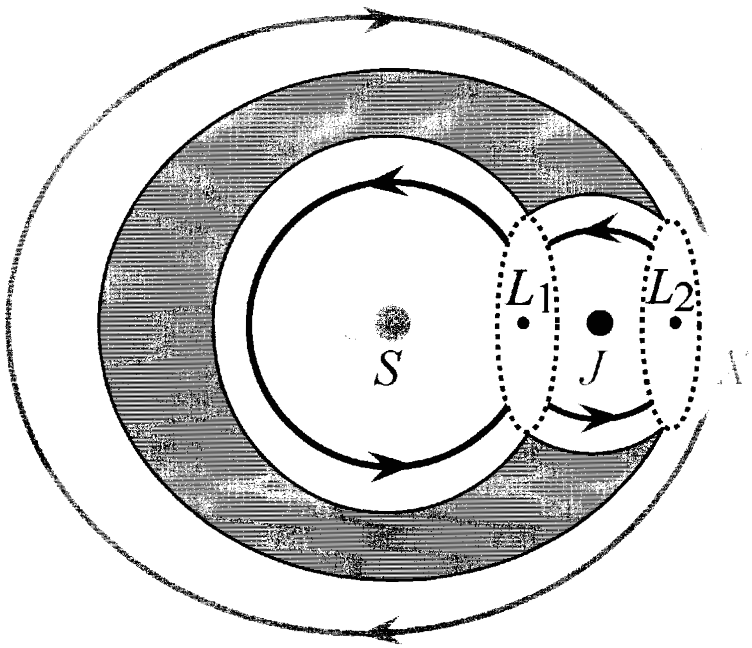


Figure 7. The Symbolic Dynamics of Transitions between the S, J, X Regions.

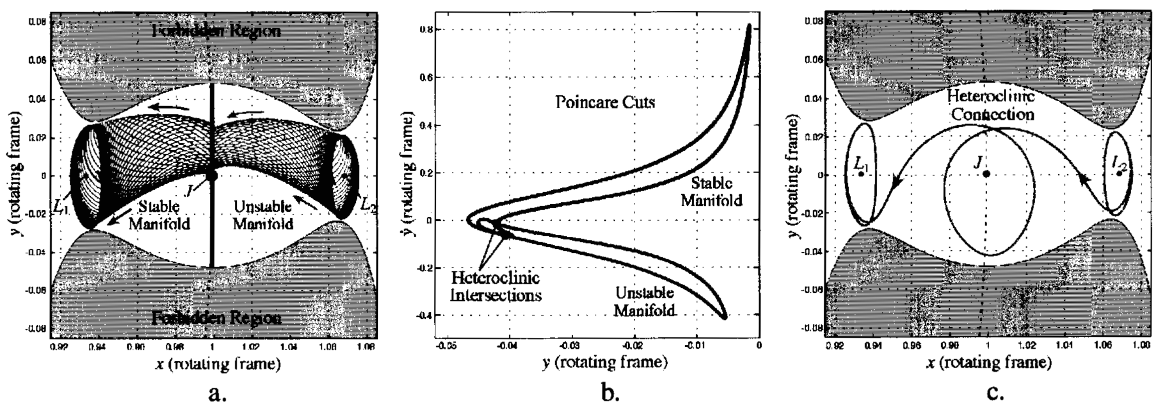
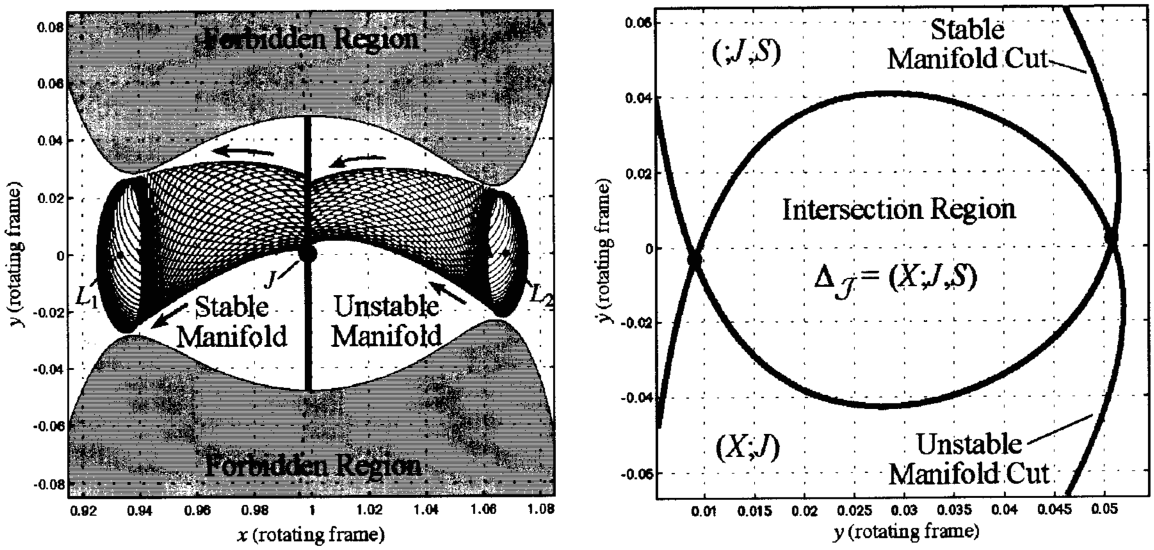


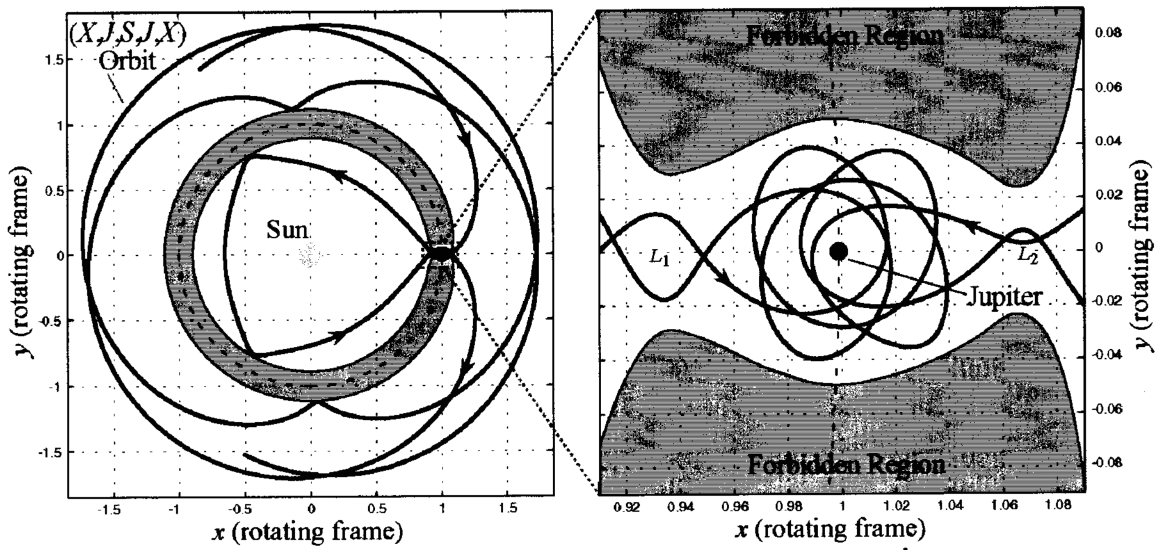
Figure 8a. The Intersection of Stable and Unstable Manifolds in the J Region.
8b. The Poincare Section of the Manifolds and Their Intersections.
8c. The Heteroclinic Orbit Generated from the Intersection.



a. b.

Figure 9a. The Intersection of the Stable and Unstable Manifolds of Periodic Orbits.

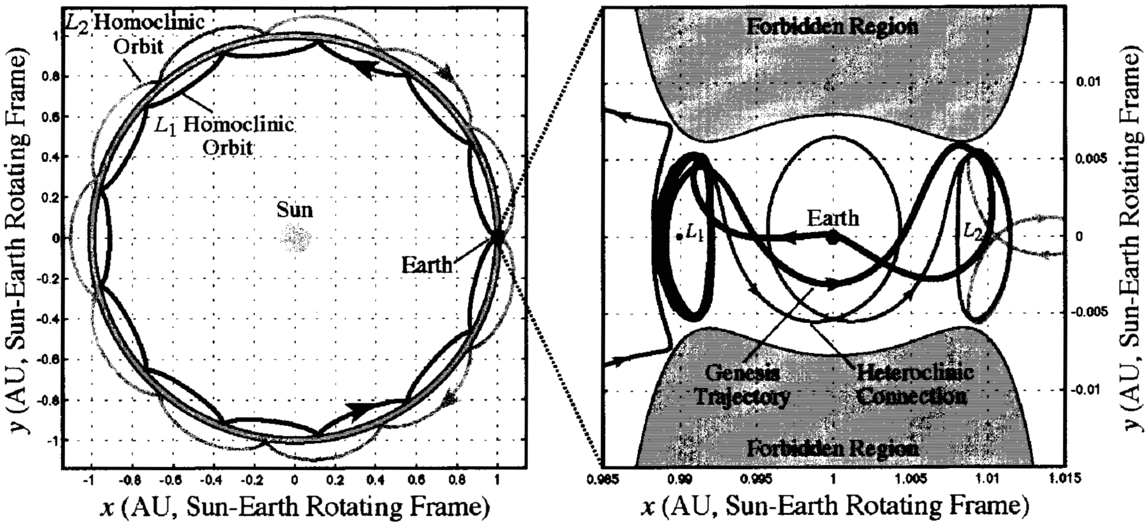
9b. Poincare Section of the Manifold Intersection. Yellow Region Depicts the (X; J, S) Sequence.



a. b.

Figure 10a. The (X, J; S, J, X) Orbit.

10b. Details of the (X, J; S, J, X) Orbit in the J Region.



a. b.

Figure 11a. The Homoclinic/Heteroclinic Chain for the Genesis Orbit.
11b. Details of the Earth Region of the Genesis Chain. Genesis Orbit in Black, Heteroclinic Orbit in Magenta.

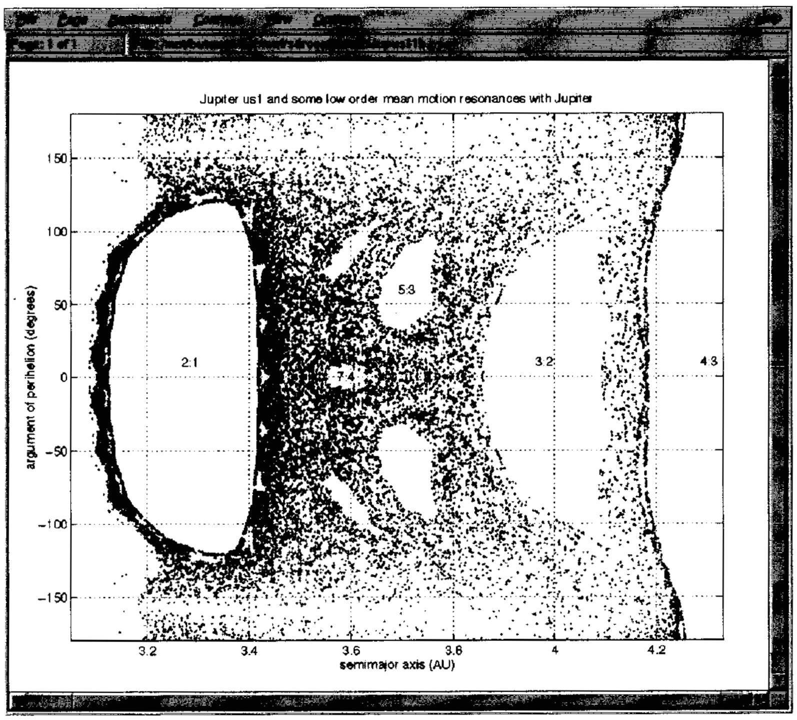


Figure 12. The Resonance Characteristics of the Jupiter L_1 Manifolds.

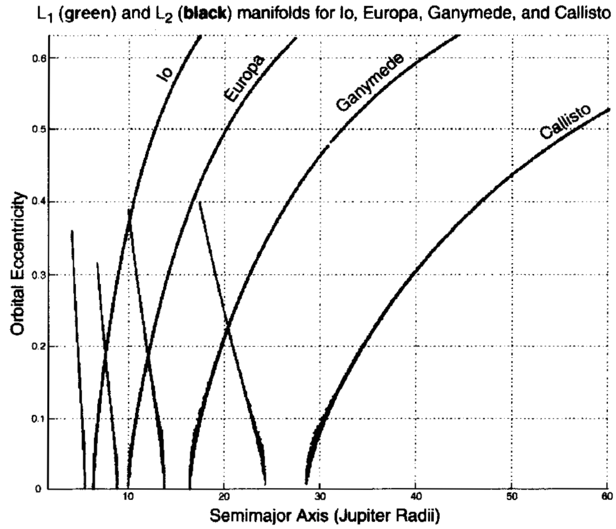


Figure 13. The Network of Interconnecting Dynamical Channels Generated by the Invariant Manifolds of L_1 and L_2 of the Jupiter Galilean Satellites.

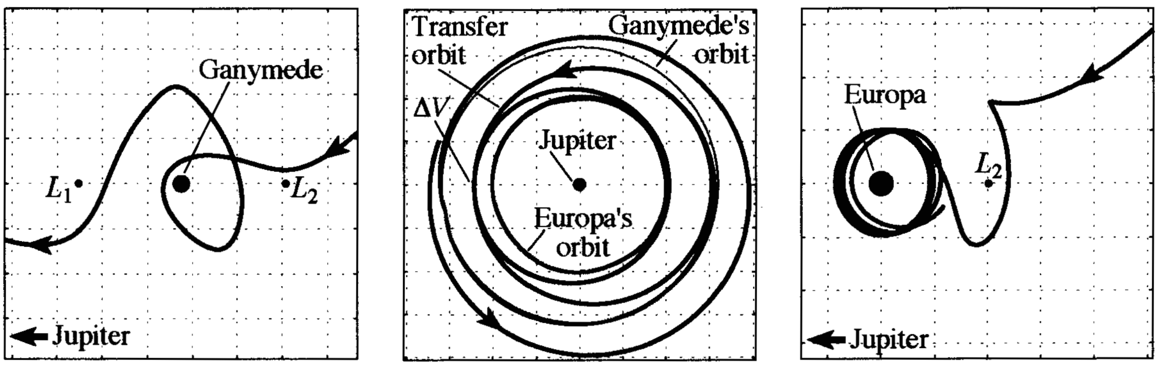


Figure 14a. One Orbit Around Ganymede in Jupiter-Ganymede Rotating Coordinates.
14b. The Transfer from Ganymede to Europa via the Invariant Manifold Intersections.
14c. Temporary Capture by Europa for 4 Orbits.