An Orbital Mechanics Primer

One: The Problem

The study of the motion of objects in space is called astrodynamics. This includes insertions of satellites into orbit around the earth, missions to the moon or Mars, and deep space interplanetary probes. Before the modern space age, scientists could learn about space only by observing natural objects in space such as stars, comets, and the planets. Planets were the only objects that could be observed with enough precision to build tables of data that could be subjected to analysis.

The first systematic observations of the planets were made by Tycho Brahe (1546 - 1601). The huge volume of data he collected was later analyzed by Johanne Kepler (1571 - 1630). Kepler was able to discern symmetry in the observed positions and published a paper in 1609 describing two principles. A third paper was published ten years later describing a third law.

Kepler's Laws

The three rules that Johanne Kepler found after studying the data for many years are:

1. The orbit of each planet is an ellipse with the sun at one focus

2. The line joining the planet to the sun sweeps out equal areas in equal times.

3. The ratio of the squares of the periods of the two planets is equal to the ratio of the cubes of the semimajor axes of their orbits.

These three laws are geometric in nature. The first says planets move in elliptical orbits. An ellipse is a common pattern in nature that has two foci. The sum of the distances from each focus to every point on the path is the same.

This is illustrated by taking a length of string and pegging both ends at the two foci, then stretching it tight with a pencil and drawing a path. Since the string does not stretch or change length, every point is equidistant from the two ends. (Notice that a circle is a special case of an ellipse, when both foci are at the same point.) Kepler found that the sun is always at one focus of the ellipse, as indicated in Figure 1.1.

The second rule Kepler found explains the speed of the planet as it moves in this elliptical path. As a planet moves around the ellipse a line from the planet to the sun sweeps out an area. The rate at which the planet moves changes according to where it is on the ellipse so that in the same amount of time it will sweep the same area. Two equal areas are crosshatched in Figure 1.2. Notice that the length of ellipse at the perimeter of the two areas is not the same length. The more distant the planet is from the sun, the slower it moves.

Notice the circle drawn dashed in the illustration. Imagine two equal areas swept out in the circle. The areas have the same length of perimeter, so the planet in orbit moves at the same speed. Any object in a circular orbit moves at a constant velocity.

Kepler's Third Law is a little subtler than the other two. It relates the dimensions of the planet's elliptical path to the planet's average rate of motion. The semimajor axis is indicated in Figure 1.3 as the distance from the geometric center of the ellipse to the most distant point on the ellipse, beyond one of the foci. The distance from the center to the nearest points on the ellipse is called the semiminor axis. The period of a planet is the total amount of time it takes to make a full orbit around the sun. Earth's period is 365 days.

Kepler compared the semimajor axis and the period of several planets and found that

$P1^2/P2^2 = s1^3/s2^3$ (1.1)

Where: P = period

s = semimajor axis

This equation is a ratio of the period squared of planets 1 and 2 to their semimajor axes cubed. The longer the semimajor axis the longer it takes a planet to move around the sun.

Vectors

This is a good time to introduce the concept of a vector, which will be important in understanding many of the ideas to be developed later in the text. A line drawn from the sun to the planet is a vector. It has a magnitude (i.e. the length) and a direction. The vector drawn in Figure 1.4 changes both length and direction according to the location of the planet on the ellipse.

When representing a vector quantity mathematically the convention here will be to use a bold font. As such, "s" is a vector with magnitude and direction and "s" has magnitude only, and is a scalar.

Notice the vector labeled "v" in Figure 1.4. It is drawn tangent to the ellipse at the point of intersection with the position vector, s. The vector v is the velocity of the planet, which is always tangent to the path. Mathematically, velocity is the rate of change of position (i.e. the more distance covered in a given length of time, the greater the velocity).

v = ds/dt (1.2)

where: v = velocity

s = distance

t = time

Velocity is the rate of change of distance, where "ds" is the change in s or displacement over a unit period of time, "dt." The greater "ds" the distance moved in a given length of time "dt" the higher the velocity. Time is a scalar quantity. It only magnitude and no direction. Also realize that speed is not the same as velocity because velocity has both direction and magnitude. The magnitude of the velocity vector is the speed.

This same relationship applies to the mathematical presentation of acceleration, which is the time rate of change of velocity.

a = dv/dt (1.3)

where: a = acceleration

v = velocity

t = time

Again, acceleration is exactly what you are accustomed to, except the mathematics used here gives it a direction as well as magnitude.

Newton's Laws

The relationships found by Kepler were geometric, they had no theoretical basis. This mathematical foundation was established by Isaac Newton (1642 - 1727) in Principia, published in 1666. Newton's first three laws created the science of motion, or mechanics.

1. Every body continues in a state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces acting upon it.

2. The change of motion is proportional to the force impressed, and is made in the direction of the straight line in which that force is impressed.

3. To every action there is always opposed an equal reaction. Or the mutual actions of two bodies upon each other are always equal and directed to contrary parts.

The second law has an elegant mathematical representation as

F = ma (1.4)

Where: F = force

m = mass

a = acceleration

Notice that mass is a scalar: it has only a magnitude. Acceleration is a vector quantity, with both magnitude and direction. When a vector and a scalar are multiplied together, the result is a vector; in this case, the force. Also, since there is only one vector on either side of the equation, the direction of both vectors must be the same.

The above three laws defined an important new concept: action because of an applied force. The constant of proportionality - the quantity that determines the magnitude of the affect on the object - is simply its mass. For a given size of force, a smaller mass will be moved faster than a large one. When vector values are used in the function, the results are just as reasonable: the object is moved in the same direction as the applied force. Conversely, an object already in motion will keep moving in the same direction until another force acts upon it.

Apply these insights to the problem of a planet orbiting the sun. The planets moved in curved paths, so there must be a force that constantly acts on the body. Otherwise once in motion it would move in a straight line only. The sun is by far the most massive object in the solar system so according to Newton's laws it is logical that it would have the predominant influence on all nearby objects. This is in fact the case: all the planets move in an orbit that is oriented with respect to the sun, which is at one focus of the ellipse.

The concept of a central force, e.g. the sun in the solar system, is common in nature. One important example is the massive, positively charged nucleus of an atom, which exerts forces that keep electrons in orbit. Electromagnetism is the central force in quantum mechanics. The central force in celestial mechanics is gravity. The importance of the central force concept in science is that similar equations apply in every case. Consequently, insights gained from the study of celestial mechanics can also provide valuable information on subatomic phenomena and vice-versa. For example, the original Bohr atomic model was based on the orbital mechanics of celestial bodies.

Universal Law of Gravitation

Newton's most famous contribution to science is the equation that defines the force of gravity.

Any two point masses attract one another with a force proportional to the product of their masses and inversely proportional to the distance between them.

The mathematical expression for this concept is

$F = G(m_1 m_2)/r^2$ (1.5)

Where: G = universal gravitational constant

m_1 , m_2 = mass of the two bodies

r = distance between the masses

Equation 1.5 gives the magnitude of the gravitational force. The direction of the force is always along the line between the masses and is directed toward the more massive body. This idea can be stated in a more formal way by establishing a three dimensional coordinate system, as shown in Figure 1.5. The x, y, and z axes are mutually perpendicular and every point in this space has three coordinates.

In most applications it is sufficient to think of only the x- and y-axes, which together define a plane. Conceptually, adding the z-axis does not complicate the presentation. So although an illustration has the three axes drawn it is OK to assume vectors are drawn in the xy plane only.

The coordinate system in Figure 1.5 is drawn with the earth at the origin. A vector is drawn from the center of the earth (which is assumed to be a point mass) to the location of a satellite. The direction of this vector is represented mathematically by a unit vector, which is a vector in the given direction that is one unit long. The unit vector is calculated by dividing the vector $\underline{r}$ by its magnitude r.

$u_r = \underline{r} / r$ (1.6)

Multiplying equation 1.5 by this unit vector gives a direction to the gravitational force.

$F = - G \times (m_1 m_2)/r^2 \times r / r = - G \times (m_1 m_2)/r^3 \times r$ (1.7)

Where the negative sign in front of G indicates the direction of the force due to gravity is opposite to the direction of the vector r, which is by convention from m_1 to m_2 as illustrated in Figure 1.5.

This format is adequate for presentation of two masses that are fixed in space. Once the force is applied, each body exerting a force on the other according to their relative masses, the bodies start to move. In order to study the resultant motion it is necessary to represent the two bodies in an unbiased coordinate system. Instead of placing the largest mass at the origin, both masses are displaced from the origin as in Figure 1.6.

Now it is possible to two separate equations for the gravity forces. One for the action of body#1 on body#2 and vice-versa. Now we have an inertial reference frame, i.e. one that is independent of the masses and which is fixed in space. This kind of coordinate system is needed in order to measure the total acceleration of each body. Now it is possible to combine equations 1.4 and 1.5

$F_1 = m_1 a = m_1 r''_1 = - G (m_1 m_2)/r^2 \times r/r$ (1.8)

$F_2 = m_2 a = m_2 r''_2 = - G \times (m_1 m_2)/r^2 \times r/r$ (1.9)

Where the acceleration is replaced by the second derivative of the displacement vector, r. That is, velocity v is the derivative of r (written as r') and acceleration is the derivative of the velocity, v'(or written as r").

Referring to Figure 1.6, it is possible to write another equation for the three displacement vectors shown.

$r'' = r''_1 + r''_2$ (1.10)

Now solve for r''_1 and r''_1 in equations 1.8 and 1.9 and substituting these values in equation 1.10. The result is independent of the vectors from the origin of the inertial reference system and is a function only of the displacement vector between the two bodies.

$r'' = G(m_1 + m_2)/r^2(\underline{r}/r) \sim G(m_1)/r^2(\underline{r}/r)$ because $m_1 \gg m_2$ (1.11)

The total mass of the two bodies is approximately equal to the mass of the larger body, e.g. the earth's mass relative to the sun's is negligible, or a satellite's mass relative to the earth. This is such a standard situation in celestial mechanics that the quantity $G(m_1)$ is replaced with a new constant, *u*, called the gravitational parameter. So, the final statement of motion is given by.

$r'' = - u(\underline{r}/r^3) = - u/r^2$ (1.12)

This says that the rate of acceleration r" between the two masses is equal to a constant divided by the magnitude of the distance between them squared, in the direction of the unit vector.

Equation 1.12 is called the two-body equation because it defines the motion of just two bodies. There were several important assumptions made to derive this function.

1. The mass of the smaller "satellite" body is negligible in calculating ¼.

2. The motion is measured relative to an inertial reference system, which is a coordinate fixed in space.

3. The two bodies are perfectly symmetrical spheres that can be represented analytically by a single point mass.

4. That is, it is possible to assume that all of the mass is concentrated at a single point in space.

5. There are no other forces acting in the system except the gravitational attraction between the two bodies.

It is extremely important to be sure, this analysis is done in an inertial reference frame. Otherwise the magnitudes of the forces calculated would be incorrect, and the conclusions drawn from the equations would be erroneous. So it is necessary to check the results. This is done by studying the motion of the two body system overall in this context.

The center of mass of a system of bodies is defined by the relationship

$r_{cm}(m_1 + m_2) = m_1 r_1 + m_2 r_2$ (1.13)

This function is best understood by comparing the values to a graph of the vectors, as in Figure 1.7. The center of mass between the two bodies is on a line drawn between them and is placed proportionately to their relative masses.

Solving equation 1.13 for the center of mass vector the result is

$r_{cm} = (m_1 r_1 + m_2 r_2)/(m_1 + m_2)$ (1.14)

Taking two successive derivatives of this function, the result is

$r''_{cm} = (m_1 r''_1 + m_2 r''_2)/(m_1 + m_2)$ (1.15)

Now it is possible to substitute for $m_1 r''_1$ and $m_2 r''_2$ and the result is that the acceleration of the center of mass is zero. This is by definition an inertial reference frame, so the issue is resolved, and the two-body equation is confirmed as a valid equation.

(NOTE: the center of mass of the earth-sun system is only 290 miles from the center of the sun - earth is 93 million miles from the sun. Could this still be within the definition of the "point mass" given the scale, making the earth-sun system an inertial reverence frame itself?)

Three-body problem

The next step in celestial mechanics is the investigation of the three-body problem. Instead of the interaction of just two bodies, the motion of three is analyzed. This extremely complex problem applies a sophisticated level of analytical methods. In fact, all of the most famous mathematicians tried their hand at some of the problems of celestial mechanic.

The difficulty in the analysis of three bodies is that the force of gravity is proportional to $(1/r^2)$. In a practical sense, as the radius decreases the force increases exponentially because you are dividing by an extremely small number. So the equations are not applicable at small distances. These distances are very important, though, because that is where most maneuvers in space occur.

A common solution to the three-body problem is to put constraints upon the analysis. Often two of the bodies are considered much more massive than the third, and all the motion is restricted to a single plane. These constraints reduce the number of unknowns in the equations of motion so that there is a valid solution (e.g. number of unknowns = number of equations). These are called restricted three-body problems and they have been the focus of most of the attention in celestial mechanics for the last century.

N-body problem

An even more complicated analysis than the three-body problem is when the interaction of many bodies is considered. For example, the affect of all the planets and the sun on a satellite moving through the solar system. This problem is so extremely complicated that virtually no attention has been devoted to it beyond setting up the most basic equations. Although mostly of interest to astrophysics, there are some general characteristics of the n-body problem that pertain to celestial mechanics when it is applied to interplanetary mission analyses.

The next chapter will discuss the calculation of escape velocity, or the velocity needed to completely escape the gravitational attraction of a body. A satellite can escape earth easily enough, but a huge amount of energy is needed to escape the solar system completely. Very few foreign objects that enter the solar system ever leave. If the gravity of the sun does not capture the massive object, then the influence of Jupiter will. (Most major planets have Jupiter at the focus and not the sun.) This massive planet serves to literally sweep the solar system of debris.

Two: The Structure

Chapter One gave mathematical and practical analysis of the motion of two bodies under the influence of gravity. This motion can be shown to be limited to conic sections (i.e. sections of a cone), of which there are three: ellipse, parabola, and hyperbola. (Circles are a special type of ellipse.) This chapter will explore the geometric facets of each of these figures because they are all quite important in celestial mechanics.

Spring-Mass System

Although the most common orbital path in nature is the ellipse, it is difficult to imagine how the force of gravity can cause such a motion. Comparing the two-body system, e.g. the sun-planet system, to a simple mechanical system may help to understand.

The two-body equation can mathematically be reduced to a Simple Harmonic Motion (SHM). This is the kind of motion exhibited by a simple spring-mass system, as illustrated in Figure 2.1. The spring elongates in direct proportion to the mass of the weight, according to the function

$F = ma = mr'' = kr \Rightarrow r'' = (m/k) \times r$ (2.1)

Where: k = spring constant

r = displacement of the spring

r'' = acceleration of the mass

If a force is applied to the mass along the axis of the spring (moving it up-down only), the mass will begin to move. Its displacement is projected to the center panel of Figure 2.1, indicating that the motion is in the exact shape of a sine wave. The reason this is called simple harmonic motion is illustrated in the right panel. If the wheel rotates at a constant rate around its center, a point on the perimeter of that wheel describes a sine wave, exactly as the spring-mass system.

Modify this simple system so that a single mass is constrained by four springs as shown in Figure 2.2. If the spring is allowed to move in two dimensions (i.e. the plane of the page), its motion will describe the shape of an ellipse.

Analytically, then, the sun has a similar affect on a planet within its gravitational field. In this instance, the gravitational force is the spring action. This action would correspond directly to our central force model, except the focus is at the center of the ellipse rather than at one of the foci.

This discrepancy can be accounted for by the addition of a small damping affect, such as shock absorbers add to the springs on a vehicle suspension.

(*NOTE: there is no damping function in the harmonic oscillator development of the two-body equation. How should the model be changed to cause the center of motion to be shifted off center, to one of the foci?)*

The ellipse

Some of the important geometric features of an ellipse are shown in Figure 2.3. The actual equation of the ellipse in a Cartesian coordinate system is

$x^2 / a^2 + y^2 / b^2 = 1$ where $a > b > 0$ (2.2)

This is the function used in analytic geometry, along with Figure 2.3. You man recall that the vertices are at the points $(-a,0)$ and $(+a,0)$ along the major axis and the foci are located at the points $(-c,0)$ and $(+c,0)$ according to the relation

$c^2 = a^2 + b^2$ (2.3)

because a, b, and c define a right triangle. In this context, the ellipse becomes a circle when $a = b$ = radius and the foci are coincident, at the center of the circle.

There are three new quantities shown on the ellipse, which is redrawn in Figured 2.4 and labeled as is common in celestial mechanics. The semiparameter (p) is the perpendicular distance from the focus to the edge of the ellipse. The periapse (r_p) is the point of closest approach to the sun and the apoapse (r_a) is the most distant point from the focus on the ellipse. The periapse and apoapse are always on the line between the foci, which is called the line of apsides. Notice that they are generic terms that have specific names depending on the object at the focus. All three of these values are vectors. Whereas the semimajor axis (a), semiminor axis (b), and half the distance between the foci (c) are scalars that describe specific dimensions of the ellipse.

Characteristics of the ellipse

The general shape of any of the conic surfaces is characterized by the eccentricity of the function. This can be defined in the following way.

Let F be a fixed point (the focus), d be a fixed line (the directrix) in a plane, and e be a fixed value. The set of all points P in the plane such that

$|PF|/|Pd| = e$ (2.4)

where | -- | denote absolute value signs. In words, the constant e is the positive ratio of the distance from the point F to the distance from the line d. It is the ratio of the distance from a point to a line such that the conic is

an ellipse for $e < 1$

a parabola for $e = 1$

a hyperbola for $e > 1$

An analogous relationship will be shown later, concerning the energy of an object moving in each of the orbital shapes. The formal definition of eccentricity is

$e = c/a$ (2.5)

or the ratio of half the distance between the foci to the semimajor axis.

There are many important functions relating a and b to the eccentricity of an ellipse.

$b^2 = a^2 (1 - e^2)$ (2.6)

$e^2 = (a^2 + b^2)/a^2$ (2.7)

Two other functions define the apoapse and periapse with respect to the eccentricity of the orbit.

$r_p = a (1-e)$ (2.8)

$r_a = a (1+e)$ (2.9)

Notice that for a circle with e = 0 the apoapse and periapse distances reduce to a, which is consistent with the geometry. Finally, there are two handy functions using the apoapse and periapse.

$2a = r_a + r_p$ (2.10)

$e = (r_a - r_p)/(r_a + r_p)$ (2.11)

These last few equations are helpful when a mission is being planned for which the apoapse and periapse distances are known and the other orbital characteristics are sought.

The hyperbola

Recall that the ellipse is the set of points on a plane, the *sum* of whose distances from two fixed points, the foci, is a constant. The hyperbola is the set of points on a plane, the *difference* of whose distances from the foci is constant. Figure 2.5 shows diagrammatically how each figure is constructed.

The equation is similar to that of an ellipse,

$x^2/a^2 - y^2/b^2 = 1$ where $a > b > 0$ (2.12) where again $c^2 = a^2 + b^2$ and the foci are at (-c,0) and (+c,0). The structure, indicated in Figure 2.6 is a little more complex. The vertices are at (-a,0) and (a,0) and are the points where the curve crosses the x-axis. The distance b is the perpendicular distance from the foci to the curve, similar to the semiparameter p on the ellipse. The curves are bounded by the two lines y = (b/a) x and y = -(b/a) x which are called asymptotes. The hyperbola approaches these lines at a great distance but never crosses them.

The parabola

The last conic surface is the set of points equidistant from a fixed point or focus, and a line called the directrix. The construction of this figure is given in Figure 2.7.

Notice the foci and the directrix are the same distance from the origin of the parabola, denoted here with the letter "p". The equation of the surface defined in this way is given by

$x^2 = 4py$ (2.13)

with the focus located at (0,p) and the directrix is the line y = -p.

Position

Another way to look at the curve of a parabola is in terms of polar coordinates. This is a slightly different coordinate system, with the same z-axes out of the page as before, but with a point designated by an angle and a radial distance in that direction. For example, the point P in Figure 2.8 is at an angle Ø and a distance r from the origin.

As before, the origin of the coordinate system is at the focus, and the directrix is a distance d away from the vertex. Re-writing the equation of the hyperbola in terms of the polar coordinates, the function is

r = ed/(1 + e cos Ø) (2.14)

which is the equation for any of the conic sections with eccentricity e. The form of this equation used in celestial mechanics for elliptical orbits is

r = p(1 + e cos Ø) (2.15)

The angle Ø is measured counterclockwise from the periapse. The distance from the focus is r and p is the semiparameter.

Try a few values in Equation 2.12 to get a feel for it. The semiparameter occurs when Ø is 90°, for which the cosine is zero, and r = p. When Ø is 0° it is the periapse, and r = p/(1+e) and when Ø is 180° it is the apoapse, and r = p/(1-e).

All other values are between these extremes, serving to define the shape of the ellipse.

This function is known as Kepler's Equation in celestial mechanics. It is one of the most important equations because it defines distance from the focus as a function of the angle. This angle is called the true anomaly, and the quantity (ed) is equivalent to the semiparameter p.

This chapter has developed some important fundamentals in the consideration of satellite orbital motion. Recall that a Cartesian coordinate system allows a mathematical analysis of position, velocity, and acceleration. The equations for radius versus angle, and the exact dimensions of the ellipse versus its parameters serve the same purpose in celestial mechanics. They create a precise reference system that can be used for a comprehensive analysis of actual motion under the influence of gravity. The only thing lacking is for time to be added to the variables. The next section describes several steps needed to accomplish this.

Eccentric Anomaly

Another common parameter in celestial mechanics is the eccentric anomaly. It is similar to the true anomaly, but it is measured from the center of the ellipse rather than from one of the foci. The eccentric anomaly, E, is related to the true anomaly, Ø, by constructing an auxiliary circle with a radius of a, the semimajor axis. A line is drawn perpendicular to the line of apses through the position of the satellite on the ellipse, and out to the auxiliary circle. The angle thus constructed is the eccentric anomaly.

An inspection of the geometry in Figure 2.10 will generate two functions for the x- and y-coordinates using the parameters of the ellipse.

$x = a \cos E - a e$ (2.16)

$y = b \sin E$ (2.17)

Inserting these values into the equation of the circle, $r^2 = x^2 + y^2$ the following function results.

$r = a (1 - e \cos E)$ (2.18)

With a little more work, it is possible to derive the mean anomaly, M. This is represented by the function

$M = n (t - t_p) = E - e \sin E$ (2.19)

Where: n = mean motion

t = time

t_p = time of periapse passage

The value t_p is the time, relative to some conventional standard, that the satellite passes through the periapse point in its orbit. The value of n is the average motion, or the total orbital period divided by the total distance traveled. That is, the period of the orbit, P = 2 À/n.

This equation is important because it is the only one that is a function of time. Consider a planet in orbit around the sun, for which you need to know the radius at a particular time. From the orbit parameters, the semimajor axis a and the time of periapse passage t_p is also known, so it is possible to calculate the mean motion from

$n^2 = \mu/a^3$ (2.20)

Now it is possible to calculate the mean anomaly M. This, in turn, is used with equation 2.19 to calculate the eccentric anomaly, which can be used in equation 2.18 to find the radius to the planet at the given time in its orbit. It is necessary to go through this complex process because the satellite moves at different speeds at different points in its orbit.

Regularization

The process of developing the mean and eccentric anomalies in order to get a solution to the two-body problem is called the process of regularization. The relationship among the known parameters is such that additional characteristics of the system must be introduced in order to solve for all the unknowns. These are artificial characteristics, as is evidence in the creation of the mean and eccentric anomalies, but they result in a function that is solvable. The best solution is a simple algebraic one, which is a closed form solution. The eccentric anomaly solution to the two-body solution is not closed, but is a transcendental solution because E cannot be solved for directly but only through numerical iteration techniques.

The process of regularization is important in the three-body problem as well. By assuming certain forms of the solution, it is possible to reduce the equations of motion to a system that can be solved. Several useful formulas have been developed to solve many restricted three-body problems, resulting in a large body of literature on the subject. Most of the results are highly theoretical, and it has been necessary to apply numerical techniques to solve the equations of motion.

(NOTE: the regularization techniques assume time is a constant, and a scalar. What if time is a vector, always constant in a given inertial reference system?)

Three: The Motion

The analysis henceforth has been a static one. Although Newton's Laws of motion were used to derive the equations, they were equations showing only the path. The force of gravity does more than exert a force and cause motion. It also determines the exact course that the movement will follow. However, depending on the magnitude and direction of the applied force, the path changes!

As you can see, even the idea of motion under the influence of gravity is not a simple one. It gets even more complicated when the natural equilibrium is disturbed by artificial forces such as a rocket injecting a satellite into orbit or thrusters causing course correction of a deep space probe.

Flight path angle

As a satellite moves around its elliptical orbit, its position changes steadily. Mathematically, its position is a vector with direction and magnitude that sweeps an equal area in an equal time. This translates into a velocity of the satellite, which is a vector that also changes constantly. Its direction is always tangent to the orbit. For a circular orbit, the tangential direction is always perpendicular to the radius of the circle to that point, or the position vector in the case of simple harmonic motion. The tangent is orthogonal to the position vector at periapse and apoapse for an elliptical path, but not elsewhere in the orbit.

The line perpendicular to the position vector at any given point on an orbit is called the local horizontal. As indicated in Figure 3.1 the angle between the local horizontal and the velocity vector tangent to the ellipse is the flight path angle. For circular orbits this angle is always zero.

The sign of the flight path angle is positive in the first half of the orbit (motion is always counterclockwise) from periapse to apoapse, and negative thereafter.

Angular momentum

There is an important relationship between velocity and position called angular momentum. Its definition is

h = r x v (3.1)

where: h = angular momentum vector

r = position vector

x = cross product

v = velocity vector

The cross product is a multiplication of vectors that is equivalent to the algebraic expression of

$h = r\ v \cos Ø_{fpa}$ (3.2)

where: r = magnitude of the position vector

v = magnitude of the velocity vector

$Ø_{fpa}$ = flight path angle

The direction of h is perpendicular to both r and v by the right hand rule. If you extend the fingers of the right hand in the direction of r and curl them toward v, the thumb points in the direction of h. Referring to Figure 3.1, the angular velocity vector points upward from the page, perpendicular to the page.

The angular velocity vector is constant for a satellite in an orbit around the sun. Both its magnitude and direction are the same at all points on the ellipse. Although the magnitude and direction of r and v and the direction of $Ø_{fpa}$ in equation 3.2 are changing throughout the orbit, together they always combine to a constant angular momentum vector.

Velocity

The simplest kind of orbit is a circular one. The radius is constant, so the velocity is constant as well. Unlike the velocity of a satellite in an elliptical orbit, the circular velocity is a function of only the radius.

$V_c = \text{sqrt}\ (\mu/r)$ (3.3)

Where μ is the gravitational parameter of the central body.

At the other extreme is the maximum velocity that a satellite can sustain and remain in orbit. This is called the escape velocity and is sqrt 2 times the circular velocity.

$V_{esc} = \text{sqrt}\ (2\mu/r)$ (3.4)

Consider a satellite in a circular or elliptical orbit. If its velocity exceeds the escape velocity, it will attain sufficient velocity to escape the gravitational influence of its central body. As indicated in Table 3.1, it will enter either a parabolic or a hyperbolic trajectory.

«Table 3.1»

Circular orbits are very important in orbital mechanics, which is the practical application of celestial mechanics to put satellites into stable orbits, and to maintain their position.

Vis-viva equation

One of the most important procedures is to alter the shape of a satellite's orbit from elliptical to circular. This process is called the circularization of the orbit. It requires a change in velocity because of the firing of thrusters. The objective is to achieve the circular velocity appropriate to the orbit radius at the time of the orbital maneuver.

The first thing is to determine the actual velocity of the satellite in its elliptical orbit. This can be determined from the Vis-viva equation.

$V^2 = \mu\ (\ 2/r - 1/a\ )$ (3.5)

Notice that for a circular orbit, a = r, and the equation is the same as equation 3.3.

The difference between the actual velocity and the planned velocity is mathematically represented as

”$V = V_{before} - V_{after}$ (3.6)

It is common to categorize this thrust maneuver as a "delta-v," which simply means that a specific variation in the velocity is accomplished.

It is most common to circularize an orbit at either the apoapse or the periapse. The resulting radius of the circular orbit will then be either r_a or r_p, respectively. The appropriate radius is substituted in Equation 3.5 for r, to find V_{before} and again in Equation 3.3 to find V_{after} . The resulting "delta-v" needed to circularize the orbit is negative if it is done at periapse, and positive at apoapse. These two directions are indicated in Figure 3.2

When circularizing an orbit along the line of apses, the direction of the "delta-v" is tangential to the orbit, which at apoapse and periapse is perpendicular to the position vector. To circularize the orbit at any other point in the orbit the velocity must still be directed perpendicular to the radius vector, but the magnitude is calculated differently.

Energy equation

Notice the column in Table 3.1 labeled "energy." The convention is that parabolic orbits are zero-energy trajectories. They are the boundary between negative energy elliptical orbits (insufficient energy to escape) and positive energy hyperbolic orbits. This is an odd concept, as it would seem that even a satellite in orbit has energy of motion. The notion of negative energy is mathematically correct, though. This is borne out by an important function in celestial mechanics called the energy integral.

$E = V^2/2 - \mu/r$ (3.7)

Where E is the energy encompassed in the satellite, with the two terms representing kinetic and potential energy, respectively. Another common expression for energy is

$E = - \mu/2a$ (3.8)

It is significant that the energy of a satellite orbit is constant, regardless of its position within that orbit. The concept of energy (even if the convention does accept "negative energy") is a useful parameter in orbital mechanics.

You will be relieved to know that all of the important equations of celestial mechanics that we will need have now been presented. There will be no more derivations. It is hoped that a few applications will enrich your understanding of these functions.

Orbital Transfer

Energy is an important consideration in orbital mechanics. Any change in an orbital profile requires thrust to change the orientation of the velocity vector. This thrust is accomplished by accelerating mass out of a jet nozzle. There is an exchange of momentum between this mass and the mass of the satellite. The mass is fuel, which must be transported at great cost in the satellite package. Therefore, it is important to design a mission so that a minimum amount of energy will be needed for orbital maneuvers.

One common scenario involves a satellite that was initially injected into a Low Earth Orbit (LEO). Once this orbit is stabilized, it is to be changed to a Geosynchronous Earth Orbit (GEO). Walter Hohmann (1880 - 1945) proved that the minimum energy transfer between two circular orbits used two tangential burns, or "delta-v's." The flight profile for such a transfer is shown in Figure 3.3.

The two circular orbits are shown with an ellipse constructed between them such that the initial orbit has a radius equal to r_p and r_a is the radius of the final orbit.

The Hohmann Transfer is done in two steps. First, a delta-V (in bold font because it has a direction and a magnitude) is applied at the perigee (i.e. the periapse of an earth-centered orbit) to inject the satellite into an elliptical orbit. Then, at the apogee of this ellipse a second delta-V is applied to circularize this orbit. The two delta-V's can be easily calculated using the functions already developed. First, the radii of the perigee and apogee are used to calculate the semimajor axis of the transfer ellipse.

$a_{trans} = (r_p + r_a)/2$ (3.9)

The initial and final velocities for the transfer trajectory are the circular velocities for the LEO and GEO, respectively.

$V_{initial} = \mathrm{sqrt}\ (\mu/r_p)$ (3.10)

$V_{final} = \mathrm{sqrt}\ (\mu/r_a)$ (3.11)

Where μ is the gravitational constant of the earth. Now the Vis-viva equation (Equation 3.5) can be used to determine the initial and final velocities on the transfer ellipse.

$V_{transA} = \mathrm{sqrt.}\ (2\mu/r_p - \mu\ /a_{trans}\ )$ (3.12)

$V_{transA} = \mathrm{sqrt}\ (\ r_p - a_{trans}\ )$ (3.13)

These four equations can now be used to calculate the required delta-V needed at the perigee and apogee.

$\text{delta-}V_A = V_{transA} - V_{initial}$ (3.14)

$\text{delta-}V_B = V_{transB} - V_{final}$ (3.15)

The total delta-V required for the transit is equal to the sum of the magnitudes of these two quantities.

Hyperbolic Trajectories

A similar analysis applies to an interplanetary transfer from, say, the orbit of Earth out to the orbit of Mars. In this instance, the sun is the central force, but the trajectory is calculated in exactly the same way (assuming the orbits of Earth and Mars are circular, which is not a bad approximation).

A new concept must be introduced when considering interplanetary travel. Strictly speaking, an object on the earth-to-mars flight path is subject to the gravitational force of the planets as well as the sun. During most of the flight, though, by far the strongest influence is exerted by the sun. As the spacecraft gets closer to mars there is a point where the planet is close enough to have the most affect on the craft.

The concept of sphere of influence was introduced to define the boundary between these two regimes. When the spacecraft reaches mar's sphere of influence the sun is "turned off" and mars is "turned on" in the analysis. The major influence is by the planet, and the sun becomes only another perturbing influence. There is no physical significance to the concept, but it is useful in mission planning studies.

Figure 3.4 shows a spacecraft approaching a planet from a great distance. Within the sphere of influence of the planet, the spacecraft moves in a hyperbolic path. Its velocity is -V(infinity) at a nominal radius of infinity. The negative sense is a function of the coordinate system chosen, which always has the approach velocity as negative.

Notice the construction includes the semimajor axis a and semiminor axis b of the hyperbola. (a is negative) The other dimensions marked in Figure 3.4 are explained as follows.

” = perpendicular distance between the asymptote and a line intersecting the focus

r_p = periapse of the approach hyperbola (closest approach)

= total change in direction between the approach velocity and the retreat velocity

Ø = the "true anomaly" of the retreat asymptote

The convention is for the semimajor axis to have a negative sign. From Table 3.1, the convention for the energy of a hyperbolic trajectory is to have a positive sense. These are consistent with the expression for energy in Equation 3.8. Substituting infinity for the radius in Equation 3.7 causes the second term in Equation 3.7 to go to zero, and the function reduces to

$E = V^2/2$ (3.16)

Which is V(infinity) or the approach/retreat velocity. Again, the sign of the energy is positive, independent of the sense of V(infinity) because the velocity is squared. To additional relationships between the geometry and the eccentricity of the hyperbola are interesting.

cos (Øinifinity) = -1/e (3.17)

sin (delta/2) = 1/e (3.18)

These functions are important in the planning of a mission. In the next chapter, it will be shown that V on approach is actually the final velocity of the Hohmann Transfer ellipse. Consequently, this is a known quantity. Another mission parameter that is known is the point of closes approach, or r_p . These two values can then be used to calculate the desired eccentricity of the approach hyperbola from the function

$E = 1 + (r_p V^2(\text{infinity}))/\mu$ (3.19)

Where µ in this case is the gravitational constant of the planet.

Four: The Trajectory

We are now ready to organize the mission to mars. All of the iterations of the two-body problem have been resolved. We also have the means to transition between the orbital conditions. The mission consists of several two-body segments, with the second body varying from the earth, to the sun, then to mars, as each becomes the most powerful influence in determining the path of the satellite. Each state is as follows:

1. Begin from a circular parking orbit around earth
2. Escape the gravity of earth on a hyperbolic trajectory
3. Make the heliocentric (sun-centered) transfer toward mars
4. Transition to mar's sphere of influence on a hyperbolic path
5. Enter a circular parking orbit around mars

Each segment is a simple two-body trajectory with the satellite as one body, and then a central force. They are conic sections, which are combined analytically in what is called the patched conic method. This is not an exact method, but is a good enough approximation for planning the mission. Once the rough mission parameters are estimated, numerical approximating methods are used to generate the final mission constraints.

Heliocentric Transfer

The overall mission parameters are input to a Hohmann Transfer trajectory. This perspective is indicated in Figure 4.1. Earth and Mars are assumed to be in circular orbits around the sun to simplify the analysis. The parameters are

r_p = average earth orbit radius

r_a = average mars orbit radius

μ = gravitational constant of the sun

The satellite path is shown by the dashed line, which is half of an ellipse.

The semimajor axis of the transfer ellipse is calculated from

$2\ a_{trans} = r_a + r_p$ (4.1)

which gives a_{trans} of approximately 188 (10^6) km. This is then used in the Vis-viva equation to calculate the velocity at the point of leaving the parking orbit around Earth.

$V_{earth} = \text{sqrt.}\ (2\mu_{sun} / r_{earth} - \mu_{sun} / a_{trans}) \sim 33$ km/sec (4.2)

This is the velocity that must be imparted to the satellite with a delta-V thrust in order for the probe to leave its circular orbit around earth and enter the transfer ellipse. However, all of this delta-V does not have to be provided by thrusters. The satellite, still in earth orbit, is moving with the earth in its heliocentric orbit already. The magnitude of this velocity is approximately

$V_{earth} = \text{sqrt}\ (\mu_{sun}/r_{earth}) \sim 28$ km/sec (4.3)

Using the earth's orbital velocity, the total velocity required is the difference between Equations 4.3 and 4.2, or 5 km/sec.

Earth Escape

The 5 km/sec is the velocity just after the satellite leaves the sphere of influence of earth. Within that sphere, as indicated in Figure 4.2, it is the hyperbolic excess velocity of earth escape, or +V(infinity).

The escape hyperbola is constructed with r_p equal to the radius of earth plus the radius of the parking orbit. This is the radius from which the delta-V is applied to project the satellite into a hyperbolic escape trajectory. Using the Energy integral, the velocity a periapse can be calculated as

V_{earth} = sqrt (V^2 + μ_{earth}/r_p) ~ 11 km/sec (4.4)

This is the total inertial delta-V needed to escape earth orbit. However, the satellite already has some velocity because it is in a circular parking orbit around earth. So this circular velocity must be deducted from Equation 4.4

delta-V = 11 km/sec - sqrt. (μ_{earth}/r_p) = 3.6 km/sec (4.5)

To review, the application of this thrust will project the satellite on the hyperbolic path indicated in Figure 4.2. At r = , the velocity on this hyperbola is the velocity required to inject the satellite into the heliocentric ellipse with Mars at the end.

Mars Approach

The satellite is now on the ellipse, and when it reaches aphelion the velocity can be calculated using the radius of mars' orbit and the semimajor axis of the transfer ellipse.

V_{mars} = sqrt ($2\mu_{sun}/r_{mars} - \mu_{sun}/a_{trans}$) ~ 22 km/sec (4.6)

This is the absolute velocity of approach to mars, the -V(infinity) on the hyperbolic trajectory once the satellite enters mars' sphere of influence, as indicated in Figure 4.3

One of the mission parameters is the height of the parking orbit desired around mars. This is used to calculate r_p, the closest point of approach. Using this value and the hyperbolic capture velocity from Equation 4.6, the required periapse velocity is then

V_{mars} = sqrt (V^2 + μ_{mars}/r_p) ~ 22 km/sec (4.4)

Which is the same value calculated in Equation 4.6.

The objective is now to put the satellite into a circular orbit around mars, so the velocity must be reduced to the circular orbit velocity

delta-V = 22 km/sec - sqrt (μ_{mars}/r_p) = 18 km/sec (4.5)

This is a very large delta-V, and the thrust required to change the satellite's velocity by that magnitude is excessive. Consequently, most mars missions will use other ways to slow the satellite's speed.

Mars Capture

You may be familiar with the idea that most man made objects in space experience a deteriorating orbit, to the extent that they eventually fall to earth. This was the fate of Skylab, and many other less renowned satellites. The reason this happens is that the earth's atmosphere, sparse though it may be at high altitudes, imposes a steady drag upon satellites. For circular orbits, the result is to reduce the height of the object until the friction becomes so great that it heats up the object, it burns and the remnants are dispersed through the atmosphere.

The affect of this drag upon an object in an elliptical orbit is a little different, as illustrated in Figure 4.4. The periapse remains fixed, but the eccentricity of the orbit slowly decreases until the orbit is circular.

Consequently, if an interplanetary mission can inject the satellite into a highly elliptical orbit (which requires a much smaller burn than to circularize the orbit) about mars, then eventually the orbit will degrade into a circular orbit. The atmosphere of mars is quite thin, so the braking action will not be great, but after many months, the objective is achieved. This process is called aerobraking.

Trajectory Correction Maneuvers

This mission planning analysis is wholly based upon the two-body problem. Different central bodies were used in each segment, but the trajectory was still simple conic sections roughly patched together. The satellite in actual transit is influenced by far more than just one body. The force of gravity from all the planets should ideally be taken into consideration, along with other perturbing effects such as solar radiation pressure, thermal affects, and so forth. These might seem like minor forces, yet we have seen that extremely small delta-v's can have a major affect upon the path of an object.

The next order of complexity possible is to investigate the "three body problem." Usually this considers the affect of two massive bodies upon a third body such as a satellite, of negligible mass by comparison. You have seen how extremely complex the structure of mathematics is for a simple two-body problem. Well, the three-body problem has no analytical solution, except in very restricted circumstances. In fact, there is a postulate of mathematics stating that it has no analytical solution.

Consequently, the presentation henceforth is just about the most sophisticated analytical solution that is possible. Any further improvement is possible only with the benefit of complex numerical techniques operating on the most sophisticated computer frames available. These refine the patched conic by adding additional forces that are exerted from other planets, solar radiation, and so forth. They do so by applying the principle of superposition. This principle states that the net affect of many different forces upon an object can be approximated by finding the affect of each one separately, then adding them together.

This process works fairly well, but it is not exact. Most deep space missions must still make many in-flight corrections, or Trajectory Correction Maneuvers (TCM's) in transit. Figure a 4.5 shows four such delta-v's that were applied on the mars Pathfinder mission. These corrections were made in order to maintain the mission on the ideal mission profile path. The critical points are the transitions from the hyperbolic trajectories to the elliptical ones, at either end of the flight path. The hyperbolas and the transfer ellipse do in fact have a uniform transition.

Again, the critical points occur at the edge of the sphere of influence of earth and mars. The forces exerted by the sun versus the nearest planet are proportional to the distance-squared. The influence by the host planet diminishes much more quickly than that of the sun, which is 333,000 more massive and whose relative distance away actually changes very little as the satellite leaves the planet's sphere of influence. Consequently, the surprise is not that the patched conic works as well as it does, but that it does not work better. Going by the relative magnitudes of the forces involved, the frequency of delta-v's required is extreme and the magnitude of divergence from the projected path is great.