MECHANICAL AND AEROSPACE ENGINEERING

# Space Flight Mechanics

## a.k.a. Astrodynamics

## MAE 589C

*Prof. R. H. Tolson*

Fall 2005

Mondays 9:00 - 11:45 A.M.

NC STATE UNIVERSITY

# Table of Contents

# Table of Contents

# List of Figures

# List of Tables

**Table** **Page**

# Chapter 1 - Coordinate Systems and Time Systems

## 1.1 Introduction

To develop an understanding and a basic description of any dynamical system, a physical model of that system must be constructed which is consistent with observations. The fundamentals of orbital mechanics, as we know them today, have evolved over centuries and have continued to require improvements in the dynamical models, coordinate systems and systems of time. The underlying theory for planetary motion has evolved from spheres rolling on spheres to precision numerical integration of the equations of motion based on general relativity. Time has evolved from using the motion of the Sun to describe the fundamental unit of time to the current use of atomic clocks to define the second. As observational accuracy has increased, models have generally increased in complexity to describe finer and finer detail.

To apply the laws of motion to a dynamical system or orbital mechanics problem, appropriate coordinate and time systems must first be selected. Most practical problems involve numerous reference frames and the transformations between them. For example, the equations of motion of a satellite of Mars are normally integrated in a system where the equator of the Earth at the beginning of year 2000 is the fundamental plane. But to include the Mars non-spherical gravitational forces (Section 5.4) requires the satellite position in the current Mars equatorial system. Planetary ephemerides (Section 2.5) are usually referred to the ecliptic, so inclusion of solar or Jovian gravitational forces require transformations between the ecliptic and the equator. The correct development of these transformations is tedious and a prime candidate for implementation errors.

Likewise, there are usually numerous time systems in a problem. Spacecraft events might be time tagged by an on board clock or tagged with the universal time that the telemetry is received at the tracking station. In the latter case, tracking station clocks must be synchronized and the time required for the telemetry signal to travel from the s/c to the tracking station must be calculated using the s/c orbit. Depending on the precision desired, this time difference might require special and general relativistic corrections. The independent variable for the equations of motion is called ephemeris time or dynamical time which is offset from universal time. By international agreement, atomic time is the basis of time and is obtained by averaging and correcting numerous atomic clocks around the world. Finally, the location of the zero or prime meridian and the equator are defined by averaging observations of specified Earth "fixed" stations. The understanding of these and other coordinate systems and time systems is fundamental to practicing orbital mechanics.

In this chapter only first order effects will be discussed. This book will also limit coverage to the classical mechanics approach, i.e. special and general relativistic effects might be mentioned but will not be included in any mathematical developments. Calculation for precise orbital mechanics and spacecraft tracking must however include many of these neglected effects. The definitive reference for precise definitions of models and transformations is the Explanatory Supplement to the Astronomical Almanac [Reference 1].

## 1.2 Coordinate Systems

The first issue that must be addressed in any dynamics problem is to define the relevant coordinate systems. To specify the complete motion of a spacecraft, a coordinate system fixed in the spacecraft at the center of mass is usually selected to specify orientation and a coordinate system fixed in some celestial body is used to specify the trajectory of the center of mass of the spacecraft. The interest here is primarily in the latter system.

**Coordinate systems** are defined by specifying

1. the location of the *origin*,
2. the orientation of the *fundamental plane*, and
3. the orientation of the *fundamental direction or line* in the fundamental plane.

The **origin** is the (0,0,0) point in a rectangular coordinate system. The **fundamental plane** passes through the origin and is specified by the orientation of the positive normal vector, usually the z-axis. The **fundamental direction** is a directed line in the fundamental plane, usually specifying the +x-axis. The origin, fundamental plane and fundamental line are defined either relative to some previously defined coordinate system or in operational terms. The definitions are usually specified in a seemingly clear statement like: “The origin is the center of mass of the Earth, the fundamental plane (x-y) is the Earth equator and the x-axis points to the vernal equinox.” Left as details are subtle issues like the fact that the center of mass of the Earth “moves” within the Earth, that the Earth is not a rigid body and the spin axis moves both in space and in the body, and that the vernal equinox is not a fixed direction. Some of these details are handled by specifying the epoch at which the orientation is defined, i.e. Earth mean equator of 2000.0 is frequently used. Further, it must be recognized that there is no fundamental inertial system to which all motion can be referred. Any system fixed in a planet, the Sun, or at the center of mass of the solar system is undergoing acceleration due to gravitational attraction from bodies inside and outside the solar system. The extent to which these accelerations are included in the dynamical model depends on accuracy requirements and is a decision left to the analyst.

Like many other fields, conventions and definitions are often abused in the literature and this abuse will continue in this text. So "the equator" is jargon for the more precise statement "the plane through the center of mass with positive normal along the spin axis." Likewise, angles should always be defined as an angular rotation about a specified axis or as the angle between two vectors. The angle between a vector and a plane (e.g. latitude) is to be interpreted as the complement of the angle between the vector and the positive normal to the plane. The angle between two planes is defined as the angle between the positive normals to each plane. The more precise definitions often offer computational convenience. For example, after checking the orthogonality of the direction cosines of the positive unit normal (usually +z axis) and the direction cosines of the fundamental direction in the plane (usually +x), the direction cosines of the +y axis can be obtained by a vector cross product. Thus, the entire transformation or rotation matrix is defined by orthogonal x and z unit vectors.

Exercise 1-1. Given vectors **a**=(3,4,-6) and **b**=(1,-3,8) in the (x,y,z) system. Define the $(\xi,\eta,\zeta)$ system such that the fundamental plane (normal to $\zeta$) contains both **a** and **b**, the $\zeta$ axis such that a right hand rotation of less than $\pi$ takes **a** into **b**, and the fundamental direction ($\xi$ axis) is along **a**.

Develop, symbolically and numerically, the 3 by 3 transformation matrix $\Phi$ from the $(\xi,\eta,\zeta)$ system to the $(x,y,z)$ system. For example, the first column of $\Phi$ is $\frac{a}{|a|} = \begin{bmatrix} 0.3841 \\ 0.5151 \\ -0.7682 \end{bmatrix}$

Common origins for coordinate systems of interest in astrodynamics include:

1. **Topocentric**: at an observer fixed to the surface of a planet,
2. **Heliocentric, Geocentric, Areocentric, Selenocentric**, etc.: at the center of mass of the Sun, Earth, Mars, Moon, etc.
3. **Barycentric**: at the center of mass of a system of bodies, i.e. the solar system, Earth-Moon system, etc.

Astronomical observations were traditionally referred to topocentric coordinates since the local vertical and the direction of the spin axis could be readily measured at the site. For dynamics problems, topocentric coordinates might be used for calculating the trajectory of a baseball or a launch vehicle. For the former case, the rotation of the Earth and the variation in gravity with altitude can be ignored because these effects are small compared the errors introduced by the uncertainty in the aerodynamic forces acting on a spinning, rough sphere. For the latter case, these effects cannot be ignored; but, gravitational attraction of the Sun and Moon might be ignored for approximate launch trajectory calculations. The decision is left to the analyst and is usually base on "back of the envelope" calculations of the order of magnitude of the effect compared to the desired accuracy.

Heliocentric, areocentric, etc. coordinates are traditionally used for calculating and specifying the orbits of both natural and artificial satellites when the major gravitational attraction is due to the body at the origin. During calculation of lunar or interplanetary trajectories, the origin is shifted from one massive body to another as the relative gravitational importance changes; however, the fundamental plane is often kept as the Earth equator at some epoch. Often in what follows only Earth geocentric systems are discussed, but the definitions and descriptions generally apply to planets and moons. Geocentric systems are either terrestrial or celestial. **Terrestrial** systems are fixed to the rotating Earth and can be topocentric, geocentric**,** or geodetic. **Celestial** systems have either the equator or the ecliptic as the fundamental plane and the vernal equinox as the fundamental direction.

### 1.2.1 Spherical trigonometry

Transformations of position and velocity vectors between coordinate systems are represented in matrix notation and developed by vector outer and inner products as mentioned above. However, the understanding of the basic concepts of spherical trigonometry is also a necessity when dealing with orbital mechanics. It is convenient to introduce the concept of the celestial sphere**.** The **celestial sphere** is a spherical surface of infinite radius. The location of the center of the celestial sphere is therefore unimportant. For example, one can think of the center as being simultaneously at the center of the Earth and Sun and observer. Any unit vector or direction can thus be represented as a point on the sphere and vice versa. For example, the Earth to Sun line and Sun to

Earth lines could be represented by two points 180 degrees apart. Two distinct points on the sphere can be connected by a **great circle** formed by the intersection on the sphere of the plane formed by the two points and the center of the sphere. If the points are not coincident or 180° apart, the great circle is unique.

The **distance** or **length** between two points on the surface is the central angle subtended by the points, which is also the shorter arc length on the great circle connecting the points. Three points, not on the same great circle, form the **vertices** of a **spherical triangle**. The three **sides** are the great circle arcs connecting each pair of vertices ($0<a,b,c<\pi$ in Figure 1-1). The length of a side of a spherical triangle is often referred to as simply the "**side**." With each vertex is associated an "**angle**" ($0 < \alpha, \beta, \gamma < \pi$) that is, the angle between the planes that form the adjacent sides. A spherical triangle has the following properties:

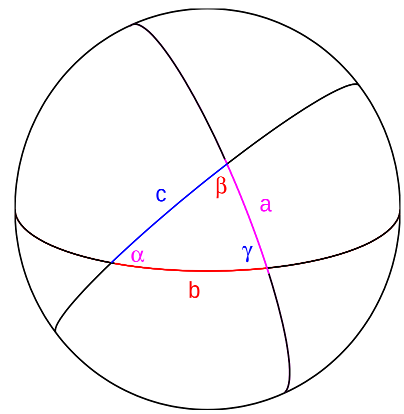


Figure 1-1. Spherical triangle

$$\pi < \alpha + \beta + \gamma < 3\pi$$
$$0 < a + b + c < 2\pi$$
$$a + b > c, \text{ etc.}$$

Exercise 1-2. Draw a spherical triangle where both $a+b+c$ is nearly zero and $\alpha+\beta+\gamma$ is nearly $\pi$. Draw a spherical triangle where both $a+b+c$ is nearly $2\pi$ and $\alpha+\beta+\gamma$ is nearly $3\pi$. Check equation (1-5) using the latter triangle.

Like plane trigonometry, spherical trigonometry relations involve four parts of the triangle. When three parts are known, the following four formulae are generally sufficient to obtain a solution for the fourth part (refer to Figure 1-1).

As in plane trigonometry there is the **law of sines**:

$$\frac{\sin a}{\sin\alpha} = \frac{\sin b}{\sin\beta} = \frac{\sin c}{\sin\gamma} \tag{1-1}$$

For spherical triangles there are two **laws of cosines.** The first is used when three sides and one angle are involved

$$\cos a = \cos b \cos c + \sin b \sin c \cos\alpha \tag{1-2}$$

and the second is used when three angles and one side are involved

$$\cos\alpha = -\cos\beta\cos\gamma + \sin\beta\sin\gamma\cos a \tag{1-3}$$

When four adjacent parts are involved, there is the **four-part formula**:

$$\cos(IS)\cos(IA) = \sin(IS)\cot(OS) - \sin(IA)\cot(OA) \tag{1-4}$$

where I and O stand for “inner” and “outer”, while A and S stand for “angle” and “side” respectively. For example, with c, $\alpha$, b and $\gamma$: $\cos(b)\cos(\alpha)=\sin(b)\cot(c)-\sin(\alpha)\cot(\gamma)$. There are three variations for each of the laws of cosines and six variations of the four-part formula. Danby [2] provides proofs of some spherical trigonometry formulae using vector analysis.

The solid angle subtended by the triangle is $\alpha+\beta+\gamma-\pi$ steradian, so if the sphere has radius R, the area of the spherical triangle is given by

$$\text{Area} = R^2(\alpha + \beta + \gamma - \pi) \tag{1-5}$$

A **right spherical triangle** has either a side or an angle of 90$^o$ and equations (1-1) to (1-4) can be reduced to two rules and **Napier's Circle**. Consider the latter case and wolog assume $\gamma$ = 90$^o$. Napier's Circle, shown in Figure 1-2, is created by putting the side opposite to the 90° angle at the top and proceeding around the triangle in the same direction to fill in the four remaining parts of the circle. The upper three parts are subtracted from 90°. Now consider any three parts of the triangle. The three parts will either be (1) “adjacent” parts, e.g. b, $\alpha$ and c in which case $\alpha$ would be called the “middle” part, or (2) two parts will be opposite the third part, e.g. b, $\alpha$ and $\beta$ and $\beta$ would be called the “opposite” part. **Napier's Rules of Circular Parts** are then:

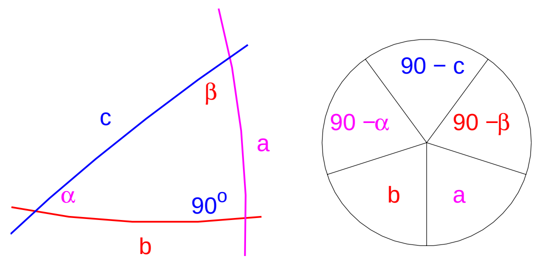



Figure 1-2. Napier’s circle.

1. The sine of the middle part equals the product of the tangents of the adjacent parts.
2. The sine of the middle part equals the product of the cosines of the opposite parts.

As stated above, the first equation is used when the three parts of interest in the triangle are adjacent, e.g. a, $\beta$ and c are related by $\cos(\beta)=\tan(a)\cot(c)$, which can be verified using equation (1-4). The second equation is used when one of the parts is opposite the other two, e.g. with b, $\alpha$, and $\beta$: $\cos(\beta)=\cos(b)\sin(\alpha)$, which can be verified using equation (1-3). Note that the quadrant is not always determined from the basic equation. Since all parts are less than $\pi$, quadrant can not be determined from sine but can be determined from tangent or cosine. Therefore, care must be exercised in determining the quadrant.

Visit http://mathworld.wolfram.com/SphericalTrigonometry.html for additional spherical trigonometry relations.

### 1.2.2 Celestial coordinate systems

The two conventional **celestial coordinate** system [1,95], projected onto the celestial sphere, are shown in Figure 1-3. The two great circles or fundamental planes of interest are the equator of the Earth and the **ecliptic**, i.e. the Earth-Moon barycenter-Sun orbital plane (often called the Earth-Sun plane). The line of intersection where the Sun, moving along the ecliptic, passes from the southern to the northern hemisphere, as seen by a geocentric observer, is called the **first point of Aries** or the **vernal equinox** and is denoted $\gamma$. The vernal equinox is the fundamental direction for celestial systems. The positive direction is from the center of the Earth to the center of the Sun at the time of the vernal equinox. This convention is one of the few remaining concepts from Ptolemy. The angle between the equator and the ecliptic is known as the **obliquity** ($\varepsilon$). The obliquity for the Earth is approximately $23.45^o$ [1,171] and changes about $0.013^o$ per century. The two intersections of the ecliptic and the equator on the celestial sphere are known as the equinoctial points. When the Sun appears to move southward through the node, it is the **autumnal equinox**. The vernal equinox occurs within a day of March 21 and the autumnal occurs within a day of September 21. At either equinox, the length of the day and night are equal at all points on the Earth and the Sun rises (sets) due east (west). When the Sun has maximum northerly declination it is **summer solstice** in the northern hemisphere and **winter solstice** in the southern hemisphere, and conversely. At summer solstice in the northern hemisphere, the longest day occurs and the Sun rises and sets at maximum northerly azimuth. Nevertheless, due to the eccentricity of the orbit of the Earth, neither the earliest sunrise nor latest sunset occurs at summer solstice. A fact that, when properly phrased, has won small wagers from non-celestial mechanicians.

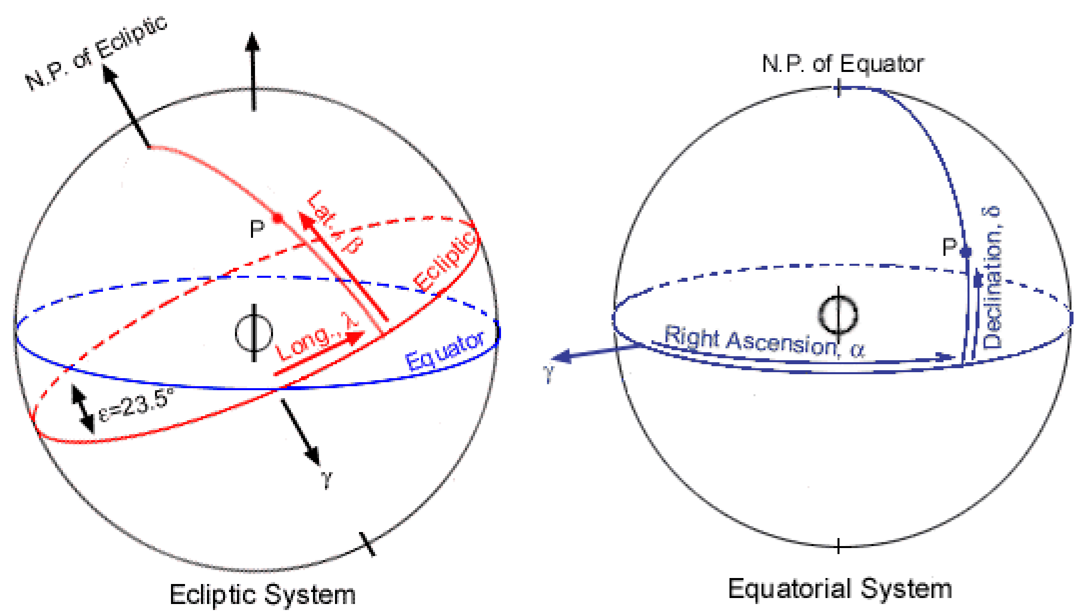



Figure 1-3. Celestial coordinate systems.

It must be recognized that neither the ecliptic nor the equator are fixed planes. Variations in the vernal equinox due to the motion of these planes are termed precession and nutation. **Precession** [1,99] is the secular component that produces a westward change in the direction of $\gamma$ that is linear with time. **Nutation** [1,109] is the quasi-periodic residual that averages to zero over many years. The **mean** equator or ecliptic refers to the position that includes only precession. The **true** equator or ecliptic refers to the position that includes both precession and nutation. The Earth equator is not fixed in space primarily due to lunar and solar gravitational torques applied to the non-spherical Earth. The **luni-solar precession** causes the mean pole of the Earth to move about 50" of arc per year and **luni-solar nutation** has an amplitude of about 9" of arc over an 18.6 year cycle. The cycle of 18.6 years is how long it takes for the orbital planes of the Earth-Moon and the Earth-Sun to return to the same relative configuration. Variations in the ecliptic are primarily due to planetary gravitational forces producing changes in the orbit of Earth-Moon barycenter about the Sun. If the equator was fixed, the **planetary precession** of the ecliptic would cause $\gamma$ to move along the equator about 12" of arc per century and the obliquity would decrease by 47" per century. To eliminate the need to consider precession and nutation in dynamics problems, the coordinate system is usually specified at some epoch, i.e. **mean equator** and **mean equinox** of 2000.0, otherwise, known as **J2000**. In this case, Earth based observations must be corrected for precession and nutation. Transformations between the J2000 coordinates and the **true** or **apparent** systems are then required [1,145].

Another plane that is use in the celestial system is the **invariant plane**. The positive normal to the invariant plane is along the total angular momentum (i.e. rotational plus orbital) of the solar system. In Newtonian mechanics, only gravitational attraction from the distant stars and unmodeled masses can cause this plane to change orientation.

Consider some point P in the geocentric reference system of Figure 1-4. The position of point P is projected onto each fundamental plane. In the equatorial system the angle from $\gamma$ to this projection is call **right ascension** $(0 \le \alpha < 2\pi)$ and the angle between the point P and the equator is called the **declination** $(-\pi/2 \le \delta \le \pi/2)$. In the ecliptic system the corresponding angles are the **celestial longitude** $(0 \le \lambda < 2\pi)$ and **celestial latitude** $(-\pi/2 \le \beta \le \pi/2)$. The "celestial" qualifier is to assure no confusion with traditional terrestrial longitude and latitude. When the context is clear, the qualifier is often omitted. "Celestial" is also sometimes replaced with "ecliptic." The rotation matrix from the ecliptic system to the equatorial system is a single rotation about the x axis by the **obliquity** $\varepsilon$. As the following example illustrates, solving spherical trigonometry problems often involves drawing numerous spherical triangle combinations until the proper combination of knowns and unknowns appears.

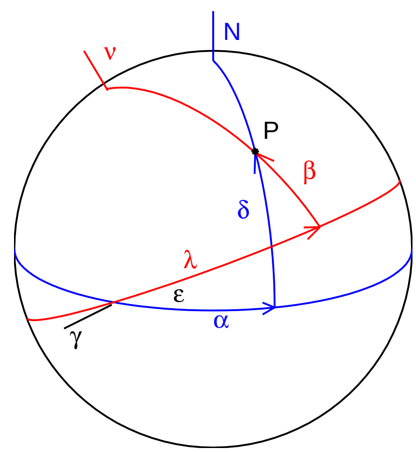



Figure 1-4. Transforming between celestial coordinate systems

Exercise 1-3. Develop the transformations from one celestial system to the other by first drawing the spherical triangle νNγ. *Notation: [xy] is the side from vertex x to vertex y and <xyz> is the angle at vertex y between sides [xy] and [yz].* Confirm that [Nγ]=[νγ]=π/2, [Nν]=ε, <Nγν>=ε, <γνN>=<γNν>=π/2. Draw the spherical triangle νNP and confirm that [NP]=π/2-δ, [νP]=π/2-β, [Nν]=ε, <PνN>=π/2–λ, and <νNP>=π/2+α. In this triangle three sides and two angles are identified, so it is possible to write α and δ as functions of only λ and β, and vice versa. Use the law of cosines and the four part formula to show

$$\sin\beta = \cos\varepsilon\sin\delta - \sin\varepsilon\cos\delta\sin\alpha \qquad \tan\lambda = \frac{\cos\varepsilon\sin\alpha + \sin\varepsilon\tan\delta}{\cos\alpha}$$
$$\sin\delta = \cos\varepsilon\sin\beta + \sin\varepsilon\cos\beta\sin\lambda \qquad \tan\alpha = \frac{\cos\varepsilon\sin\lambda - \sin\varepsilon\tan\beta}{\cos\lambda} \tag{1-6}$$

As a check, (1) notice that the second line can be obtained from the first line by switching equatorial and ecliptic variables and replacing $\varepsilon$ with $-\varepsilon$, (2) if $\varepsilon=0$ then $\alpha=\lambda$ and $\beta=\delta$, (3) if $\alpha=\pi/2$ ($3\pi/2$) then $\lambda=\pi/2$ ($3\pi/2$) and $\beta=\delta+\varepsilon$ ($\delta-\varepsilon$). Since parts of spherical triangles are by definition less than $\pi$, determine the validity of the transformations when $\alpha$ or $\lambda$ are greater than $\pi$. Develop tests or new equations to eliminate any quadrant ambiguities or singularities.

### 1.2.3 Terrestrial coordinate systems

Astrodynamics problems are generally framed in either the ecliptic or equatorial celestial coordinate system. The locations of observers, receivers, transmitters, and observation targets are usually specified in one of the terrestrial coordinate systems. A **terrestrial coordinate system** [1,199] is "fixed" in the rotating Earth and is either geocentric or topocentric. Transformations between terrestrial and celestial coordinates are an essential part of orbital mechanics problems involving Earth based observations. These transformations are defined by the **physical ephemeris** (Section 1.4), that is, the definition of the pole location and the rotational orientation of the Earth. Precise definitions must include elastic deviations [1,237] in the solid Earth, plate tectonics [1,249], motion of the spin axis in the Earth [1,238], and numerous other effects. The largest of these effects is polar motion which produces deviations between the instantaneous and mean spin axis of order 10 meters. Pole location is determined by numerous observation stations and published by international agreement. Irregularities in the rotational rate of the Earth can change the length of the day by a few milli-seconds over time scales of interest for orbital mechanics and astronomy problems. One $\text{milli}-\text{second} \approx 0.46$ meters in longitude at the equator. Rotational variations are also monitored and included in the definition of universal time to be discussed later. Specific effect to be included depend on the desired accuracy and the choice is left to the analyst.

The **fundamental terrestrial coordinate** system has the origin at the center of mass and the equator as the fundamental plane. The intersection of the reference meridian with the equator is the fundamental direction. The **origin, the equator, and reference meridian** [1,223] **are defined operationally by measurements made at a number of "fixed" stations on the surface**. In the past, the prime meridian was the Greenwich meridian and was defined by the center of a plaque at Greenwich. The phrase "reference meridian" is used to clearly distinguish the fundamental difference in definitions. Nevertheless, the reference meridian is often referred to as the Greenwich meridian, and that practice will be used herein. For remote solid planets, prime meridians are still defined by easily observed sharp surface features. An observer's **local meridian** is defined by the plane through the observer that also contains the spin axis of the Earth. An observer's **longitude** ($\lambda$) is the angle between the reference meridian and the local meridian, more precisely referred to as "**terrestrial longitude**." Since the spin axis moves in the Earth, an observers true longitude deviates from the mean longitude.

Latitude is specified as either geodetic latitude ($\phi$) or geocentric latitude ($\phi'$) (Figure 1-5.) **Geocentric latitude**, often called **latitude**, is the angle between the equator and the observer. In the **geocentric system**, the location of a point is specified by the **radius** from the center of the Earth, **geocentric latitude** and **longitude**. To satisfy the right hand rule convention for rotations about the pole, longitude should be measured east; but, is often measured west. To be safe, always specify the convention. For example, $75^{o}$W = $285^{o}$E longitude. **Colatitude** is the angle between the position vector and the normal to the equator and is unambiguous, but latitude is sometimes specified by using a sign convention e.g. -37.5°=37.5°S. Also note that geocentric latitude is often denoted by $\phi$, i.e. the "prime" is omitted when the meaning is clear.

**Geodetic coordinates** are generally limited to points near the surface of the Earth. **Geodetic latitude** is the angle between the local vertical and the equator. The **local vertical** is determined by the local "gravity" force which is the combination of gravity and a centrifugal contribution due to rotation. An equipotential surface for the two terms is nearly an ellipsoid of revolution. Hence it is convenient to define a **reference ellipsoid** (spheroid) for the mean equipotential surface of the Earth which is approximately the **mean sea level.** This ellipsoid, which is symmetric about the equator and has rotational symmetry about the pole, is defined by the **equatorial radius** (a) and the **flattening** (f). The polar radius is given by b =a(1-f). Reference values [1,223] are a=6378137m and 1/f=298.25722. Figure 1-5 shows a cross section of the reference ellipsoid with greatly exaggerated flattening. For the figure, it is assumed that the cross section contains the x-axis, so the equation of the elliptical cross-section is

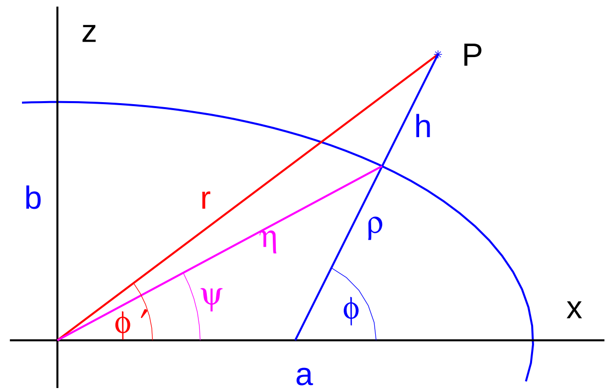



Figure 1-5. Geodetic and geocentric latitude

$$f(x,z) = \frac{x^2}{a^2} + \frac{z^2}{b^2} - 1 = 0 \qquad (1\text{-}7)$$

Exercise 1-4. The gradient of f, $\nabla \mathbf{f}$, when evaluated at a point on the surface of the reference ellipsoid (h=0 in Figure 1-5) is a vector normal to the surface (B-2) and pointing outward. From this vector develop the following relationship between geodetic and geocentric latitude [3,78]

$$\tan\phi = \frac{\tan\phi'}{(1-f)^2} = \left(\frac{a}{b}\right)^2 \tan\phi' \text{ (EE)} \qquad (1\text{-}8)$$

Geodetic longitude and geocentric longitude are equal. For a point above the reference ellipsoid, the **geodetic altitude** (h in Figure 1-5) is defined as the closest distance to the reference ellipsoid and the **geodetic latitude** ($\phi$ in Figure 1-5) is defined as the angle between the normal to the ellipsoid and the equator at this closest point. Points with the same geodetic altitude are nearly on the same equipotential surface. Global atmospheric models, often used to calculate drag on a satellite Section 5.5.1, generally assume hydrostatic equilibrium and geodetic altitude is often an independent variable in the model. Determining the geodetic altitude and latitude from geocentric position $(\rho, \phi')$ is not straight forward but can be done exactly [1,206] or approximately. The basis of most methods is to start with the location of a point P in both systems and introduce the auxiliary angle $\psi$ leading to the three equations

$$\begin{aligned} x = r\cos\phi' &= \eta\cos\psi + h\cos\phi \\ z = r\sin\phi' &= \eta\sin\psi + h\sin\phi \\ \tan\psi &= (1-f)^2\tan\phi \end{aligned} \tag{1-9}$$

These equations can be solved by Newton-Raphson iteration or successive substitution for geodetic altitude and geodetic latitude and $\psi$, noting that $\eta$ is a known function of $\psi$. Or $\psi$ can be eliminated by noting that $\eta\cos\psi = aC\cos\phi$ and $\eta\sin\psi = a(1-f)^2 C\sin\phi = aS\sin\phi$, where $C = [\cos^2\phi + (1-f)^2\sin^2\phi]^{-1/2}$ and $S=(1-f)^2*C$. The transformation from geodetic to geocentric is obtained directly from equations (1-9).

**Topocentric coordinate** systems are also of interest. The origin of the system is fixed on the surface of the planet. For example, the location of a satellite relative to a ground based tracking system utilizes this frame. The fundamental plane for topocentric coordinates is either normal to the geocentric radius (geocentric topocentric) or tangent to the reference ellipsoid (geodetic topocentric). In both cases the fundamental plane is called the **horizon**. The points directly overhead and directly beneath the origin or observer are called the **zenith** and the **nadir**, respectively. The plane, formed by zenith and the north pole, is called the **meridian** and where it intersects the horizon is usually the fundamental line. Coordinates of points in the topocentric frame are specified by **range** ($\rho$), **azimuth** (A) and **elevation** (a). Range is the distance from the origin to the point. Azimuth is specified as either east or west of North. Sometimes "east" is taken as the fundamental direction and azimuth is given as north or south of east. It is best to be explicit, e.g. 32.5° E of N. The elevation angle is zero for points on the horizon and 90° for points at zenith. The **zenith angle** is the complement of the elevation. Be aware that astronomers call elevation "altitude" and in Sonnet 116, Shakespeare calls it "height."

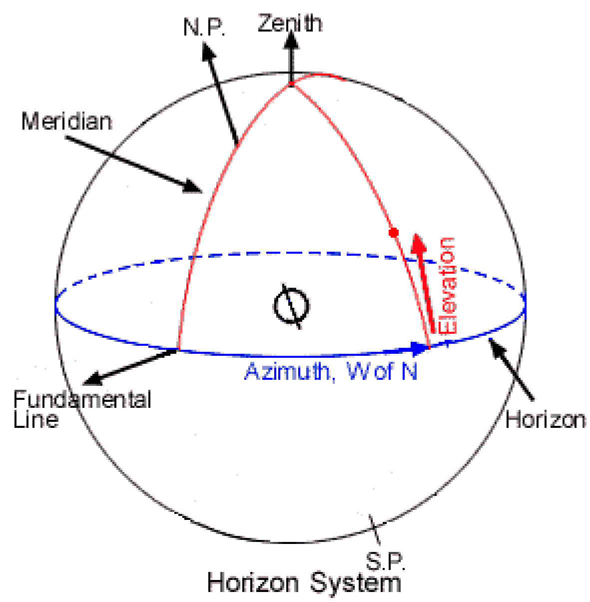



Figure 1-6. Topocentric coordinate system.

## 1.3 Time Systems

The above descriptions of the various spatial coordinate systems may initially leave the reader in a confused state of mind. The situation is only slightly better for time systems. St. Augustine wrote, "If you don't ask me what time is, then I know; but, if you ask me, I don't know." After reading this section you will probably appreciate his statement. Nevertheless, understanding the relevant time systems is essential to any orbital mechanics application. Time is generally thought of as a linearly increasing scalar. So a **time system** is defined by: (1) the unit of constant **duration** and (2) the zero value or **epoch** i.e. like any well defined line, a slope and an intercept. In celestial mechanics problems, three time systems are used. These are

> **Universal time** or civil clock time which accounts for both the rotation and orbital motion of the Earth with respect to the Sun and is generally the independent variable for measurements.
>
> **Sidereal time** which is a measure of the rotation of the Earth relative to the vernal equinox and locates the Earth based observer in the celestial coordinate system.
>
> **Ephemeris time** or **dynamical time** which is the independent variable for orbit calculations and locates spacecraft, planets, etc. in the celestial coordinate system.

All of these times are related to the atomic time which is fundamental by international agreement. The following description of these four systems are short versions of and in some places approximations to the detailed descriptions given in Reference 1.

### 1.3.1 Atomic time

The fundamental unit of **atomic time** [1,40] is the Système International second or **SI second**. This is defined as the duration of 9,192,631,770 periods of the radiation from the transition between two levels of the ground state of the cesium-133 atom. This duration was adopted to be consistent with ephemeris time (Section 1.3.3). Within our current understanding of physics, the SI second is a fixed number. However, the definition is operational so measurements are required to determine atomic time. Further, relativistic corrections must be made to these Earth based measurements. The time standard that most closely follows the definition is the International Atomic Time or Temps Atomique International (**TAI**). TAI is supplied by the Bureau International des Poids et Measures in Sèveres, France. To obtain TAI an intermediate time scale is determined by combining data from a number of high-precision atomic standard clocks. This intermediate time scale is available in real time. After the fact, corrections are made for known effects to achieve a time as close as possible to atomic time. This adjusted time scale is published as the TAI.

### 1.3.2 Dynamical time

The independent variable in the equations of planetary motion [1,41] is **dynamical time**. Theories of relativity states that this value depends upon the reference coordinate system as well as the particular theory. To reduce periodic contributions and produce a nearly constant duration, the origin of the reference system is taken at the barycenter of the solar system and is called **barycentric dynamical time** (**TDB**). On the other hand, **terrestrial dynamical time** (**TDT**) is a theoretical time scale constructed from apparent geocentric ephemerides of bodies in the solar system. Dynamic time in other systems are then available by transformations and conversely.

With respect to TAI: TDT=TAI+32.184$^s$. The offset between TAI and TDT is set to provide continuity with ephemeris time which was the independent variable in the EOM until dynamical time was introduced. The offset is equal to the estimate of the difference between ET and TAI when TDT was introduced.

### 1.3.3 Ephemeris time

**Ephemeris time** (ET) was developed as the independent variable for Newton's laws of motion and theory of gravitation. ET is a uniform time scale to depict observations of bodies in the solar system. There are three different forms of ET (ET0, ET1, and ET2), each based on more complex models of lunar motion. There is no detectable rate difference between ET and UTC (Section 1.3.6), but the epoch difference is updated with leap seconds. Although ephemeris time has been formally replaced by dynamical time the two are often used synonymously.

### 1.3.4 Julian date

The **Julian date** is simply a means of continuously counting the number of days from an epoch sufficiently far in the past to precede the historical record of astronomical observations. This continuous count is done with **Julian day numbers**. The first Julian day number (0) is defined as Greenwich mean noon on January 1, 4713 BC in the Julian proleptic calendar or Nov. 24, 4713 BC in the modern calendar. Note: **JD starts at noon**! Julian dates can be expressed in UT or dynamical time. For celestial reference coordinate systems, the epoch is defined in TDB. Thus, **J2000.0** is 2000 January 1.5 TDB, which is JD 2451545.0 TDB. A **Julian century** is 36525 days. For convenience, the **modified Julian date** (MJD) was defined as the value of JD minus 2400000.5. **MJD starts at midnight!** There are a number of formula for converting from a Gregorian date to Julian date. The issue is of course how to handle the leap years. A year is a leap year if it not a century year and is divisible by 4. If the year is a century year, it is a leap year only if it is divisible by 400 (e.g. 1600 and 2000 are leap years while 1700, 1800, 1900 are not).The algorithm[3,61]

$$\begin{gathered} A = \langle Y/100 \rangle \qquad B = 2 - A + \langle A/4 \rangle \\ JD = \langle 365.25(Y + 4716) \rangle + \langle 30.6001(M + 1) \rangle + D + B - 1524.5 \end{gathered} \tag{1-10}$$

is valid for all positive JD, Y is Gregorian year, M is month (3 to 14), D is day of the month including any fractional part. Note that month is increased by 12 for January and February and year is decreased by one. The 0.5 on the last term accounts for JD starting at noon and the 0.0001 addition is to assure largest integer operator performance. The symbol $\langle x \rangle$ is the largest integer operator which is the largest integer less than or equal to x. In MATLAB use the "floor" operator. The following inverse transformation is valid only for JD>2299161.

$$\begin{gathered} z = \langle JD + 0.5 \rangle \qquad f = JD + 0.5 - z \\ a = \langle (z - 1867216.25)/36524.25 \rangle \qquad b = z + a - \langle a/4 \rangle + 1525 \\ c = \langle (b - 122.1)/365.25 \rangle \qquad d = \langle 365.25c \rangle \qquad e = \langle (b - d)/30.6001 \rangle \\ D = b - d - \langle 30.6001e \rangle + f \\ \text{If } e < 14, \quad M = e - 1 \quad \text{else} \quad M = e - 13 \\ \text{If } M > 2, \quad Y = c - 4716 \quad \text{else} \quad Y = c - 4715 \end{gathered} \tag{1-11}$$

The more general algorithm is given in [3,63]. It is common to give time as year, day of year and seconds into the day (YY,DOY,S). Given month (M) and day of the month (D), the day number (DOY) can be calculated from

$$\begin{aligned} \mathrm{DOY} &= \langle (275\mathrm{M})/9 \rangle - 2\langle (\mathrm{M}+9)/12 \rangle + \mathrm{D} - 30 \qquad \mathrm{non-leap-year} \\ \mathrm{DOY} &= \langle (275\mathrm{M})/9 \rangle - \langle (\mathrm{M}+9)/12 \rangle + \mathrm{D} - 30 \qquad \mathrm{leap-year} \end{aligned} \tag{1-12}$$

Other useful algorithms on time transformations can be found in Reference 3 (Note that this reference is not without typographical error.)

### 1.3.5 Sidereal time

**Sidereal time** [1,48] is defined as the hour angle of the vernal equinox. An observers **hour angle** is the angle from the vernal equinox, measured eastward, to the observers meridian. As such, sidereal time is a measure of the diurnal rotation of the Earth. **Apparent sidereal time** is the hour angle of the true equinox as defined by the true equator and true ecliptic of date, i.e. apparent sidereal time includes the nutation in $\gamma$ and therefore includes **periodic inequalities**. **Mean sidereal time** is the hour angle of the mean equinox and includes only the precession of $\gamma$ and therefore only **secular inequalities**. Apparent sidereal time minus mean sidereal time is the **equation of the equinoxes**.

Sidereal time on the Greenwich meridian is called **Greenwich sidereal time** (Section 1.3.8). **Local sidereal time** is the Greenwich sidereal time added to the local east longitude. Sidereal time is traditionally stated in hours, minutes, and seconds with one hour corresponding to fifteen degrees of rotation relative to the vernal equinox. A **sidereal day** is defined as the period of consecutive passes of the equinox. Due to precession in $\gamma$ the mean sidereal day is shorter than the period of rotation of the Earth by about 0.0084 seconds. The sidereal day begins with the first transit of the vernal equinox (**sidereal noon)** and ends with the second transit.

### 1.3.6 Universal time

The basis for all civil time-keeping [1,50] is known as Universal Time (UT). **Universal Time** is derived from the mean diurnal motion of the Sun and incorporates the rotational and orbital motion of the Earth with respect to the Sun. **UT0** is determined directly from measurements of fixed stellar radio sources and depends on the observer location. UT0 accounts for variations in pole location and non-uniform rotation. These effects must be considered in precision orbital mechanics problems requiring tracking station location or any other geo-location to a few meters.

**UT1** is obtained when UT0 is corrected for the shift in longitude caused by the motion of the pole relative to the surface of the Earth. UT1 is global because it is based on a mean pole location. UT1 is not a uniform time scale due to variations in the rotational rate of the Earth. The current definition of UT1 was created to fulfill the following conditions [1,51]:

1. UT1 is proportional to the angle of rotation of the Earth in space, reckoned around the true position of the rotation axis,
2. the rate of UT1 is chosen so that the day of UT1 is close to the mean duration of the solar day, and

3. the phase of UT1 is chosen so that $12^h$ UT1 corresponds approximately with the instant when the Sun crosses the Greenwich meridian.

**UT2** is UT1 corrected for variations in the Earth rotation rate and has a uniform rate but not the same as TAI. The final form of universal time, Coordinated Universal Time (**UTC**) is used by broadcast time services. UTC differs from TAI by an integer number of seconds and is kept within 0.9 seconds of UT1 by the use of leap seconds generally at the end of June or December. By definition, UTC and TAI run at the same rate.

Figure 1-7 shows a graphical description of the relevant times scales. The reference is TAI on the x-axis. TAI vs. TAI has a slope of 1 and intercept at zero. TDT or ephemeris time has an observed slope of one but an offset of 32.184 seconds. UT1 on average has a slope less than 1 because of a rate difference between the rotation of the Earth and TAI. UTC is kept within 0.9 seconds of UT1 by introduction of a “leap second” (http://hpiers.obspm.fr/webiers/general/earthor/utc/UTC.html) as appropriate. There was a “leap second” at the beginning of 1999 and during the rest of the year there was a constant offset between TDT and UTC of 64.184 seconds. **This is a critical number in celestial mechanics problems because the ephemerides are integrated in TDT (ephemeris time) and observations are time tagged in UTC.**

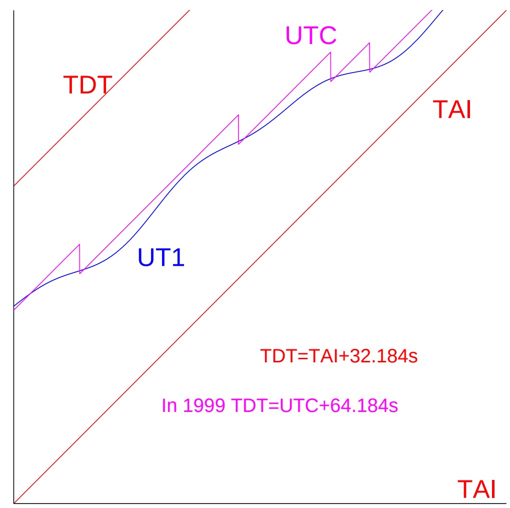



Figure 1-7. Summary of time scales.

### 1.3.7 UT1, UTC and Pole Location for 1998

Recall that UT1 is determined operationally by satellite tracking, lunar laser ranging and very long baseline radio interferometry (VLBI) data averaged over the globe. VLBI provides the most accurate measurements with an accuracy of about 0.00005 seconds when averaged over one day [1, 62]. Figure 1-8 shows the measured difference between UT1 and UTC during 1998. Since the difference between UT1 and UTC did not approach 0.9 seconds, there was no “leap second” in 1998. Recent leap seconds were Jan. 1, 1999, July 1, 1997, Jan. 1, 1996. There have been 22 leap seconds from 1972 through 2005 or about every 18 months. Leap seconds were included every year since 1972 except for 1984, 1986, 1987, 1989, 1995 and 1998-2005. A leap second will occur at the beginning of 2006.

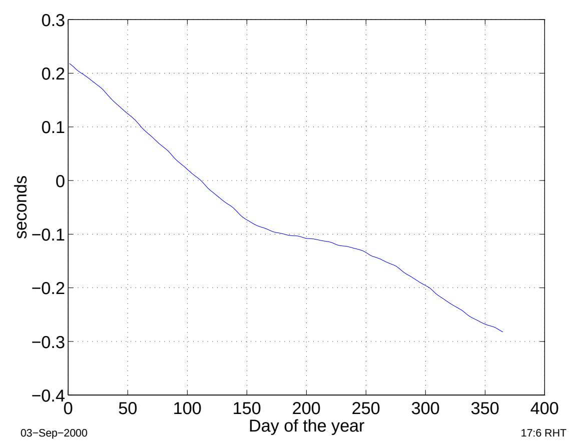



Figure 1-8. UT1 - UTC for 1998.

Figure 1-9 shows the pole location during 1998. The difference between UT0 and UT1 is primarily due to the location of the pole, but continental drift and other small effects contribute to the time difference. If the surface features of the Earth at some instant of time are considered fixed, then the instantaneous axis of rotation of this surface defines the pole of the Earth. Even if there were no external torques on the Earth, the pole would move with respect the surface due to natural precession of a torque free, rotationally symmetric, rigid body, as discussed in most dynamics book. For the Earth this motion is called the Chandler wobble after the person who provided an explanation in 1891 of the observed variation in latitude. If there were no external torques, the amplitude of this motion would be expected to damp to zero over long periods of time due to friction in the oceans and the elastic Earth. However continual excitation is provided by external torques due to lunar and solar gravity and internal motions of the Earth due to seasonal variations in atmospheric and ocean mass distributions, earthquakes, and any other phenomena that change planetary moments of inertia. The data were taken from http://hpiers.obspm.fr/webiers/general/earthor/polmot/PMOT.html.

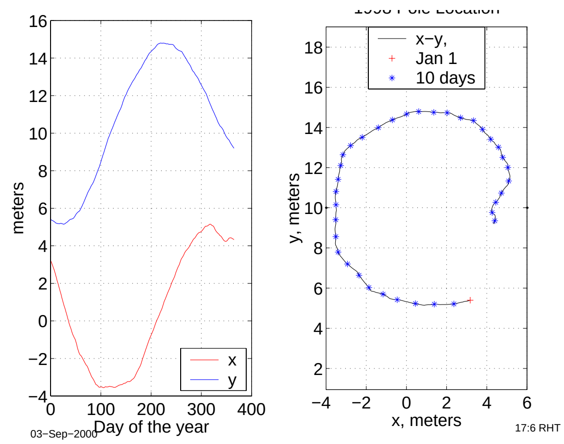


Figure 1-9. Pole location for 1998, x is along the Greenwich meridian, y is west.

Exercise 1-5. Visit http://maia.usno.navy.mil and http://hpiers.obspm.fr. Write a two page paper on what you discovered about UT1-UTC, pole location and anything else related to this chapter. E.g. make a plot of x vs. y pole location in meters and/or UT1-UTC for the last full year. Check the figures above.

### 1.3.8 Greenwich and local mean sidereal time

**Greenwich mean sidereal time** (**GMST**) is the angle between the Greenwich meridian and the mean vernal equinox and would be sensitive to the same variations in rotation as UT1. Due to the three conditions above, there is now a defined relationship between the two at midnight:

$$\begin{aligned} \text{GMST1-0hUT1} &= 24110.54841\text{sec} + 8640184.812866T + 0.093104T^2 - 6.2 \times 10^{-6}T^3 \\ &= 100.4606184° + 36000.77005T + 0.00038793T^2 - 2.6 \times 10^{-7}T^3 \end{aligned} \quad (1\text{-}13)$$

where in the first equation the coefficients are in seconds of time, T = d/36525 is the number of Julian centuries and d {±0.5, ±1.5...} is the number of days of UT elapsed since Julian Date 2451545.0 UT1 (Jan. 1, 12h, 2000). In the second form of the equation, GMST1 has been converted to degrees of rotation from $\gamma$ to the Greenwich meridian by multiplying the first equation by 360/86400. The “1” at the end of GMST notates that the value is based on UT1 and correction to UT0 may be required for local observers.

Some consequences of these relationship are discussed in Reference 1. It is readily shown that the ratio of mean sidereal rate to UT1 rate is **r'=1.002737909350795** plus secular terms that affect the eleventh decimal place per century. From this it follows that the **mean sidereal day** is $23^h56^m04^s.090524$ of UT.

The information for calculating the Greenwich and local mean sidereal time is now in place. Refer to Figure 1-7 where “GM” represents the Greenwich meridian and “LM” represents the local meridian.

Given year, month, day and clock time or UTC in hours h, e.g. h=17.5678395 hrs:

1. Correct UTC to UT1 if necessary. This is only required if geographic locations to better than a few meters are required.
2. Use equation (1-11) to calculate the Julian date (JD) at 0 hours.

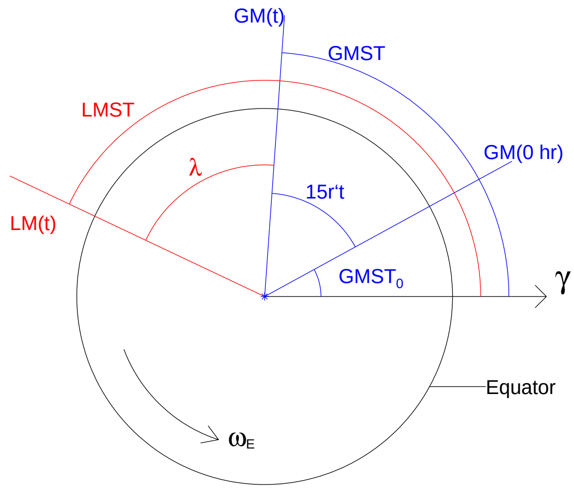



Figure 1-10. Local and Greenwich sidereal time

3. Calculate $T = \dfrac{JD - 2451545.0}{36525} =$ Julian centuries from noon of Jan. 1, 2000.
4. Calculate the $GMST1_0°$ at 0 hours using equation (1-13).
5. Calculate the change in GMST1 since 0 hours using $\Delta GMST1° = 15r'h$.
6. Finally $GMST1° = GMST_0° + \Delta GMST1°$ gives the angle from the mean equinox to the Greenwich mean meridian.

The **local mean sidereal time** at east longitude λ is $MST(\lambda°) = GMST1° + \lambda°$; thereby providing the final information necessary to transform between the mean celestial and geocentric, Earth fixed coordinate systems. Transformations to J2000 [1,99] would have to include precession (50.290966"/year) and nutation.

## 1.4 Physical Ephemerides

Using the celestial and terrestrial coordinate systems and the time systems defined above, it is possible to transform from terrestrial coordinates at any time to celestial coordinates. It is necessary to have the same capability for the Moon, the planets, planetary moons, and other bodies in the solar system. The means for making such transformation is through the physical ephemeris. The **physical ephemeris** of a body defines a body centered, body fixed coordinate system relative to a celestial system. The parameters that define this system are called the **rotational elements** for the body and consist of the direction of the rotational pole, the rotation rate, and the location of the prime meridian at some epoch. The prime meridian for bodies without solid visible surfaces is generally taken as the central meridian as seen by a geocentric observer at some epoch. For such bodies, the rotational rate may be latitude dependent. For bodies with solid surfaces, the prime meridian is generally associated with a sharp surface feature, e.g. the central peak in a crater. The table below gives the physical ephemeris for the terrestrial planets. The pole location is given by the right ascension and declination referred to J2000. Locations with asterisks indicate that a secular variation is included in the model and the secular precession values are given in [1,401]. The location of the prime meridian $W = W_0 + \dot{W}_0 t$ is measured along the planet equator from the line of intersection of the planet equator with the Earth equator of J2000, with W=0 corresponding to the point where the equator of the planet passes into the northern hemisphere of the Earth, i.e. the ascending node of the planet equator with respect to the Earth equator.

Table 1-1. Physical Ephemerides for the Terrestrial Planets

| | **Pole location** | | **Prime meridian** | | **Equatorial radius, km** | **Flattening f** |
|---|---|---|---|---|---|---|
| | **$\alpha_o$, deg** | **$\delta_o$, deg** | **$W_o$, deg** | **$\dot{W}_o$, °/day** | | |
| Mercury | 281.01* | 61.45* | 329.71 | 6.1385025 | 2439.7 | 0.00 |
| Venus | 272.72 | 67.15 | 160.26 | -1.4813596 | 6501.9 | 0.00 |
| Mars | 317.68* | 52.89* | 176.868 | 350.8919830 | 3397. | 0.0065 |

## 1.5 Problems

1-1. Given the latitude and longitude of two cities, derive one equation from spherical trigonometry to determine the distance between the cities. Apply the equation to find the distance (km) from Washington, DC to Chicago, IL.

1-2. Given the latitude and longitude of two cities, derive one equation from spherical trigonometry to determine the departure azimuth for a great circle flight between the cities. Apply the equation to find the azimuth (deg) from Chicago to DC.

1-3. For an observer at latitude $\phi$, determine the azimuth of sunrise and sunset in terms of the declination of the Sun using one equation from spherical trigonometry. Assume the Sun is an infinite distance from the Earth. Evaluate the equation for an observer at Washington, DC and at both equinoxes and solstices.

1-4. Why is the sidereal day shorter than the day based on universal time? Develop two approximate estimates of r' (Section 1.3.6) based on (1) the number of days in the year and (2) the number of days in a Julian century. Compare with precise value.

1-5. Does the vernal equinox occur earlier or later in a leap year than in non-leap years? Why?

1-6. Perform a manual calculation to determine the mean sidereal time of 1300 EST July 4, this year in Washington DC.

## 1.6 Astronautics Toolbox

1. Write a function [$\phi$', r]=Geod2Geoc(h, $\phi$, a, f, ichk) that converts geodetic altitude and latitude to geocentric radius and latitude. The output variables and the first two input variables are n by 1 arrays while a and f are scalars. Test the function at a variety of altitudes using the supplied test program (TestGeod2Geoc) and the supplied function Geoc2Geod.

2. Write a procedure JD=Ymd2JD(ymdhms, ichk) where ymdhms=[YYYY,MM,DD,HH,MM,SS.SSS], an n by 6 array and JD an n by 1 output array. Write a test procedure (TestYmd2JD) that test the function using the supplied JD2Ymd.
3. Write a procedure Gmst=Ymd2Gmst(ymdhms,ichk), where ymdhm=[YYYY,MM,DD,HH,MM,SS.SSS] as above and Greenwich mean sidereal time as n by 1 array output. Include range testing, fatal errors, and test procedure. Use Ymd2JD.m if necessary.
4. Write a procedure R=J2k2Planet(JD, $\alpha_o$, $\delta_o$, $W_o$, $\dot{W}_o$, ichk), where JD is Julian day (1 by 1), and the other parameters are taken from Table 1-1. R is the 3 by 3 rotation matrix (B-1) from J2000 to the planet equator, prime meridian system.

## 1.7 References

In the text, references are identified in general by [n; m; etc.]. However, if a specific page is identified then the notation is [reference number, page], i.e. [3,47]. The reference number is generally a hyperlink.


1. Seidelmann, K. **Explanatory Supplement to the Astronomical Almanac**, University Science Books, Mill Valley, CA, 1992.
2. Danby, J.M.A., **Fundamentals of Celestial Mechanics**, Willmann-Bell, Richmond, Va., 1989.
3. Meeus, John, **Astronomical Algorithms**, Willmann-Bell, Inc., Richmond, Va., 1991.


## 1.8 Naval Academy Pledge response to being asked for the time

**xxxx-1977**

Sir, I am greatly embarrassed and deeply humiliated, but due to circumstances beyond my control, the inner workings and hidden mechanisms of my chronometer are in such in accord with the great sidereal movement by which time is recorded, that I can not with any degree of certainty state the exact time. However, I will estimate that the current eastern standard time is 31 minutes, 44 seconds, and 1 tick past the hour of 14.

**1977-**

Sir, I am greatly embarrassed and deeply humiliated, but due to circumstances beyond my control, the printed circuit board in my chronometer is in such a chaotic state compared to the atomic clocks by which time is recorded, that I can not with any degree of certainty state the exact time. However, I will estimate that the current eastern standard time is 1F minutes, 2C seconds, and 1 bit past the hour of E.

# Chapter 2 - N-Body Problem

## 2.1 Introduction

The description of the motion of a system of n bodies due to their mutual gravitational attraction is the fundamental problem in orbital mechanics. Applications range from the stability of the solar system to the formation of galaxies. No closed form solution exists for the general n-body problem when n is greater than two. However, it was shown by Lagrange that solutions do exist for special cases of the three-body problem, all of which require that the motion of the bodies takes place in the same plane. These special cases will be discussed in Chapter 4.

Before discussing the n-body problem, some of the fundamental principles of mechanics will be reviewed. Among these are: Newton's laws of motion, the concepts of work and energy, and the concept of angular momentum. It is also useful to be aware of the theory of general relativity equations of motion for the n-body problem.

## 2.2 Newtonian Mechanics

Newton formalized the physical laws which determine the dynamics of massive bodies. Based on earlier work of Galileo, Kepler and others, he established three laws of mechanics and one for gravitational attraction. These laws were adequate to predict the dynamical motion of the planets and terrestrial objects for hundreds of years. Only after significant increases in observational precision was it necessary to seek modifications. The laws were formulated for particles and integration over the volume is required for application to finite bodies. The laws are only valid in an inertial frame. It is often said that such a frame is at "rest" or moving with constant velocity. Such a statement implies the existence of some absolute frame to which such motion can be referred. It might be said that an inertial system is at rest or moving with uniform velocity relative to the fixed stars. The problem has now been transformed to defining the "fixed" stars. An equally acceptable definition is to say a system is **inertial** if Newton's laws of motion are valid in that system. For practical applications, the analyst can pick a system moving through space with origin at the center of mass of the solar system or perhaps one whose origin coincides with the center of the Earth. It may even be reasonable to regard a system of coordinates attached to the Earth's surface as inertial, provided the accelerations resulting from the translation and rotation of the system are negligible compared with the acceleration of the body under consideration. The choice of coordinate systems is purely an issue of the accuracy desired in the prediction of the motion, there is no system that is exact and the choice is left to the analyst.

### 2.2.1 Laws of motion

Newton's three classic laws can be stated as follows:

**First Law:** *If there are no forces acting on a particle, the particle will move in a straight line with constant velocity.*

In Newtonian mechanics, a **particle** is a point mass. This is a fundamental concept requiring no further definition. In practice, a finite body can be shown to behave like a particle in some cases or can be considered to be a particle if the physical dimensions are small compared to the distance to other bodies. In Newtonian mechanics, **force** and **position** are also fundamental notions requiring no definition. Denote by **f** the force vector and by **v** the velocity (i.e. the time derivative of position) in an inertial space.

**Second Law:** *A particle acted upon by a force moves so that the force is equal to the mass times the time rate of change of the velocity.*

In equation form

$$\mathbf{f} = \mathbf{m}\frac{d\mathbf{v}}{dt} = m\mathbf{a} \tag{2-1}$$

If "force" is fundamental, then "mass" is just the proportionally constant and conversely. Force and mass can not be defined independently. The first law, which Galileo discovered by rolling spheres down incline planes, is a special case of the second law.

**Third Law:** *When two particles exert forces upon one another, the forces are of equal magnitude and in opposite directions.*

This law is often called the law of **action and reaction**. Denoting by $\mathbf{f}_{ij}$ the force exerted by particle j upon particle i, then the law states $\mathbf{f}_{ij} = -\mathbf{f}_{ji}$.

### 2.2.2 Law of universal gravitation

Newton's law of universal gravitation was based on Kepler's laws of planetary motion [Section 3.2] and is the force model required to satisfy the condition that the orbital period is proportional to the 3/2 power of the semi major axis. The **universal gravitation** law is stated as: *two particles of mass* $m_1$ *and* $m_2$ *attract each other with a force along the line joining the two particles and with a magnitude proportional to the product of the masses and inversely proportional to the square of the distance between the particles.* Following the notation above, this is mathematically

$$\mathbf{f}_{ij} = \frac{-Gm_im_j\mathbf{e}_{ij}}{r_{ij}^2} = \frac{-Gm_im_j\mathbf{r}_{ij}}{r_{ij}^3} \tag{2-2}$$

Where $\mathbf{e}_{ij}$ and $\mathbf{r}_{ij}$ are the unit vector and position vector from $m_j$ to $m_i$, and G is the **universal gravitational constant** ($6.672\times10^{-11}$ $m^3/kg/s^2$). The law is known as the **inverse square law**. In practice G is almost never used because observations determine the product GM to much higher precision than G can be determined. For the Earth GM=398600.5 $km^3/s^2$.

Though the inverse square law is formulated for point masses, it also holds for bodies with a spherically symmetric distribution of density. It is sufficient to show that the attraction between an exterior particle of mass $m_1$ and a thin spherical shell of constant density and mass $m_2$ satisfies the inverse square law.

As shown in Figure 2-1, the mass $m_1$ is at a distance R from the shell center. From equation (2-2), $m_1$ is attracted with a force

$$\mathbf{f} = Gm_1 \int \frac{\mathbf{r}}{r^3} dm_2$$

Due to the symmetry of the problem, all components of **f** normal to the line between $m_1$ and the center of the shell will cancel, so the direction of the resulting force is along the line between $m_1$ and the center of $m_2$. By integration it can be shown that the magnitude is

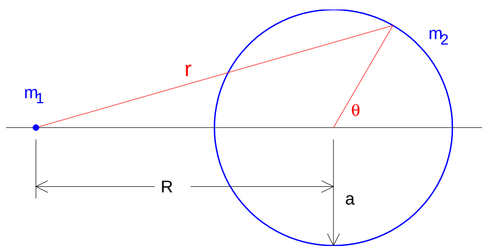



Figure 2-1. Attraction of a homogeneous spherical shell

$$f = \frac{Gm_1m_2}{R^2} \tag{2-3}$$

Thus, the inverse square law holds for homogeneous spheres as well as for particles. This is one example where a finite mass can be considered a particle.

Exercise 2-1. Fill in the steps to verify equation (2-3).

### 2.2.3 Kinetic and potential energy

The concepts of **work**, **kinetic energy** and **potential energy** are also important in celestial mechanics. **Work** is a scalar quantity defined as the line integral of force along a particular path

$$W_{12} = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{f} \cdot d\mathbf{r} \tag{2-4}$$

between positions $\mathbf{r}_1$ and $\mathbf{r}_2$. Note that the definition has nothing to do with dynamics, particles or time, and implicit in the definition is the assumption that **f** depends only on position. The concept can be extended to work performed by the force that produces the motion of a particle by using Newton's second law to eliminate **f** in the integral and set d**r**=**v**dt. In this case, it can be shown that the work done on particle m is just the change in kinetic energy between the end points

$$W_{12} = \frac{m}{2}(v_2^2 - v_1^2) \tag{2-5}$$

Exercise 2-2. Starting with equation (2-4), verify equation (2-5) showing that the change in kinetic energy is the work done by the external forces.

A force $\mathbf{f}(\mathbf{r})$ is **conservative** if $\oint \mathbf{f} \cdot d\mathbf{r} = 0$ when taken about any closed path. For such a force, the work defined by equation (2-4) is independent of the path from $\mathbf{r}_1$ and $\mathbf{r}_2$. The work and change in kinetic energy are functions of the end points only.

Exercise 2-3. Using only the definition, show that for a conservative force the work performed in moving between two points is independent of the path taken to get from one point to the other.

The concept of **potential energy** at a point $V(\mathbf{r})$ can now be introduced as the negative of the work done by a conservative force in going from a reference point $\mathbf{r}_o$ to an arbitrary point $\mathbf{r}$

$$V(\mathbf{r}) = -\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{f} \cdot d\mathbf{r} + V(\mathbf{r}_0) \tag{2-6}$$

Within an additive constant, a scalar **potential** can therefore be uniquely associated with every point in space. So that the work done in going from $\mathbf{r}_1$ to $\mathbf{r}_2$ given by equation (2-5) can be expressed in terms of the potential as

$$W_{12} = V(\mathbf{r}_1) - V(\mathbf{r}_2)$$

This implies that

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{f} \cdot d\mathbf{r} = -\int_{\mathbf{r}_1}^{\mathbf{r}_2} dV$$

or since $r_1$ and $r_2$ are arbitrary

$$\mathbf{f} \cdot d\mathbf{r} = -dV$$

which permits the force to be expressed as the negative gradient (B-2) of the potential

$$\mathbf{f} = -\nabla V(\mathbf{r}) \tag{2-7}$$

The force is called a **conservative force** because total energy is conserved during the motion due to such a force. That is, if along the trajectory $\mathbf{r}_i$ and $\mathbf{v}_i$ are the position and velocity at time $t_i$ and $T_i$ and $V_i$ are the corresponding kinetic and potential energies, then

$$T_1 + V_1 = T_2 + V_2$$

Given $\mathbf{f}(x,y,z)$, the operational test to determine if $\mathbf{f}$ is conservative comes from the theorem (B-2): $\mathbf{f}$ is conservative if and only if $\nabla \times \mathbf{f} = \mathbf{0}$. The potential function can then be determined in principle from equation (2-7) and boundary conditions. For example, the force $\mathbf{f}(x,y,z) = (-2xy, z^2-x^2, 2yz) = -2xy\mathbf{e}_x + (z^2-x^2)\mathbf{e}_y + 2yz\mathbf{e}_z$ is derivable from a potential function since

$\nabla \times \mathbf{f} = \left(\frac{\partial f_z}{\partial y} - \frac{\partial f_y}{\partial z}\right)\mathbf{e}_x + \left(\frac{\partial f_x}{\partial z} - \frac{\partial f_z}{\partial x}\right)\mathbf{e}_y + \left(\frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y}\right)\mathbf{e}_z$ is readily shown to vanish. From equation (2-7) $\frac{\partial V}{\partial x} = 2xy$ implies that $V=x^2y + g(y,z)$, where g is any arbitrary function of y and z. Also $\frac{\partial V}{\partial y} = x^2 - z^2$ so that $\frac{\partial g}{\partial y} = -z^2$ or $g = -yz^2 + h(z)$, where h(z) is an arbitrary function of z. Now $V=x^2y - yz^2 + h(z)$. This form of V identically satisfies the final term in the gradient $\frac{\partial V}{\partial z} = -2yz$ so that h is a constant and wolog select the boundary condition so that h=0.

Exercise 2-4. Show that the force $\mathbf{f}(x,y,z)=(2xz/r^4, 2yz/r^4, 2z^2/r^4-1/r^2)$ is conservative and derive the potential function. As usual $r^2 = x^2 + y^2 + z^2$.

### 2.2.4 Linear and angular momentum

The **linear momentum** of a particle is the mass times the velocity

$$\mathbf{p}=m\mathbf{v}. \tag{2-8}$$

Newton's second law is often stated as the time rate of change of linear momentum equals the force. The **moment of momentum** or **angular momentum** is another important concept in mechanics. For a particle of mass m at position **r** and with linear momentum **p**=m**v**. The **angular momentum** about the origin from which **r** is measured is defined by

$$\mathbf{h} = \mathbf{r} \times m\mathbf{v} \tag{2-9}$$

Often no distinction is make between angular momentum and **specific angular momentum**, i.e. angular momentum per unit mass. Even though two satellites of the Earth can have significantly different masses, if they are in the same orbit they will be said to have the same angular momentum. This is done because, as will be seen in Section 3.3, the orbital characteristics are determined by the sum of the masses of the Earth and the satellite and the latter is generally of negligible mass compared to the former.

## 2.3 Equations of Motion

While the two body problem discussed in Chapter 3 can be applied to many cases, and has the advantage of having a closed form solution, certain problems cannot be modeled with sufficient accurately using this assumption, and must be solved as a general system of n bodies. Consider a system of n bodies where each body is either spherical symmetry or sufficiently far from other bodies that each can be regarded as a point mass. It will be assumed that the only forces acting upon the system are due to the mutual Newtonian gravitational attraction. Let the mass and the position of each body in the system be denoted by mass $m_i$ and $\mathbf{r}_i$ and the vector from mass $m_j$ to mass $m_i$ by $\mathbf{r}_{ij}=\mathbf{r}_i-\mathbf{r}_j$. From Newton's second law and law of universal gravitation, for each mass

$$m_i\ddot{\mathbf{r}}_i = -G\sum_{j\neq i}^{n}\frac{m_i m_j}{r_{ij}^3}\mathbf{r}_{ij} \qquad i = 1\ldots n \tag{2-10}$$

where the notation j≠i means the sum over all values of j excluding i.

## 2.4 Integrals of the Motion

Equations (2-10) are a set of n second order, non-linear, coupled, ordinary differential equations and the solution will require 6n independent constants of integration. The constants of integration are usually determined from the n position vectors and the n velocity vectors at some epoch. Of the 6n required integrals of the motion only 10 are known. The relationships between these integrals and the physical assumptions are

1. No external forces and "action and reaction" assures conservation of total linear momentum,
2. Mutual force along the line between bodies and no external torques assures conservation of total angular momentum, and
3. Conservative force field and no external energy transfer assures conservation of total system mechanical energy.

Each of these conservation laws will now be demonstrated from the equations of motion (EOM).

### 2.4.1 Conservation of total linear momentum

The location of the **center of mass** (CM) of the system is given by

$$\mathbf{R} = \frac{1}{M}\sum_{i=1}^{n} m_i\mathbf{r_i}$$

where M is the total mass. Since this equation is true for any time, it can be differentiated with respect to time to get the EOM of the CM location. Performing this operation and eliminating the $\ddot{\mathbf{r}}_\mathbf{i}$ using equation (2-10) yields $\ddot{\mathbf{R}} = 0$. Integrating twice yields

$$\mathbf{R}(t) = \mathbf{V_0}(t-t_0) + \mathbf{R_0}$$

Thus the CM of the system or barycenter of the system moves with constant linear velocity $\mathbf{V}_o$. The vectors $\mathbf{V}_o$ and $\mathbf{R}_o$ represent six integrals of the equations of motion.

The **total linear momentum**, **P** is defined as the sum of all the individual linear momenta, i.e.

$\mathbf{P} \equiv \sum_{i=1}^{n} m_i\mathbf{v_i} = M\dot{\mathbf{R}} = M\mathbf{V_0}$. So that the **total system linear momentum is conserved**.

Exercise 2-5. Fill in the steps to verify the motion of the center of mass and the conservation of linear momentum.

### 2.4.2 Conservation of total angular momentum

The **total angular momentum, H** is the sum of the individual angular momenta. As usual, to test for conservation of angular momentum, each EOM is pre-crossed with the corresponding position vector and the result is summed over all bodies. Since $\frac{d}{dt}(\mathbf{r} \times \dot{\mathbf{r}}) = \mathbf{r} \times \ddot{\mathbf{r}}$, $\mathbf{r_{ij}} = -\mathbf{r_{ji}}$ and $\mathbf{r_i} \times \mathbf{r_{ij}} = \mathbf{r_j} \times \mathbf{r_i}$, the result is

$$\frac{d\mathbf{H}}{dt} = \frac{d}{dt}\left[\sum_{i=1}^{n} \mathbf{r_i} \times m_i \dot{\mathbf{r}}_\mathbf{i}\right] = 0 \tag{2-11}$$

which represents the statement of the **conservation of the total angular momentum**. The three constant components of **H** constitute three additional integrals of the motion. Equation (2-11) implies that both the magnitude and direction of vector **H** are constant. The constant direction of **H** can be used to define a plane through the center of mass of the system. This plane was called the **invariant plane** by Laplace. For the solar system the invariant plane is inclined at about 1º35' with respect to the ecliptic, between the orbital planes of the two most massive planets Jupiter and Saturn. Except for the attraction of mass outside the solar system, the invariant plane is inertial in Newtonian mechanics.

Exercise 2-6. Fill in the steps to verify equation (2-11).

### 2.4.3 Conservation of energy

The **total mechanical energy**, E is the sum of the individual kinetic and potential energies. To test for conservation of energy, use the equations of motion to form an expression that looks like the rate at which the forces are doing work. This is accomplished by forming the dot product of **v** with each EOM and summing the results over all bodies to get the total rate at which work is being done. Since $r\frac{dr}{dt} = \frac{1}{2}\frac{d}{dt}(r^2) = \frac{1}{2}\frac{d}{dt}(\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}) = \dot{\mathbf{r}} \cdot \ddot{\mathbf{r}}$ , the sum can be written as

$$\frac{d}{dt}\left(\frac{1}{2}\sum_{i=1}^{n} m_i(\dot{\mathbf{r}}_\mathbf{i} \cdot \dot{\mathbf{r}}_\mathbf{i}) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j \neq i}^{n}\frac{Gm_i m_j}{r_{ij}}\right) = \frac{d}{dt}(T + V) = 0 \tag{2-12}$$

Where the first term is recognized as the **total kinetic energy** and the second term as the **total potential energy**. Thus **total mechanical energy is conserved**, i.e. T+V=E=constant.

Exercise 2-7. Fill in the steps to verify equation (2-12).

The energy integral is the tenth and last known constant of the motion for the n-body problem. The ten constants are **Vo**, **Ro, H,** and E. The existence of additional integrals has been investigated extensively and summarized in Moulton [1]. Brun has shown that the 10 known integrals are the only independent integrals in rectangular coordinates that are algebraic and Poincare has shown that if the orbital elements [Section 3.3] are used as generalized coordinates there are no independent uniform transcendental integrals. Jacobi has also shown that if all but two of the integrals exist it is always possible to get the last two integrals. Hence, the ten integrals above plus Jacobi's theorem assures a solution of the two body problem.

## 2.5 Planetary Ephemerides

An **ephemeris** is a tabular representation of the motion of some body. A **planetary ephemeris** is a tabulation of the motion or trajectory of a planet and a **satellite ephemeris** is a tabulation of the motion or orbit of a satellite. Prior to modern computer technology and the GPS constellation, the planetary ephemerides were published annually as a listing of the position of the planets throughout the year. This information was very useful to astronomers and navigators. These tables were of sufficient accuracy for most optical observations. Early tables were recorded on paper and generally included the position and difference tables at uniform time intervals. Lagrange or other interpolation polynomials were used to determine intermediate positions. Current ephemerides are in a similar format but of course recorded on computer compatible media. Ephemerides can be defined with various levels of accuracy. The most accurate ephemerides are generated using the equations of motion from general relativity. Less accurate ephemerides are generated by omitting various terms from the equations of motion.

### 2.5.1 General relativity

The theory of general relativity is thought to completely describe the gravitational interaction of bodies [2]. However, the interaction is so complicated that even the one body problem, i.e. a particle with negligible mass being attracted by a body of finite mass, has not been solved. Approximations must be made to even write the equations of motion. For a body in the solar system the equation of motion is given to order $1/c^2$ as:

$$\ddot{\mathbf{r}}_{\mathbf{i}} = -\frac{\sum_{j\neq i}\mu_j \mathbf{r}_{\mathbf{ij}}}{r_{ij}^3}\left[1 - \frac{4}{c^2}\sum_{k\neq i}\frac{\mu_k}{r_{ij}} - \frac{1}{c^2}\sum_{k\neq j}\frac{\mu_k}{r_{jk}} + \frac{v_i^2}{c^2} + 2\frac{v_j^2}{c^2} - 4\frac{\mathbf{v}_{\mathbf{i}}\cdot\mathbf{v}_{\mathbf{j}}}{c^2} + \frac{3}{2c^2}\left(\frac{\mathbf{r}_{\mathbf{ij}}\cdot\mathbf{v}_{\mathbf{j}}}{r_{ij}}\right) - \frac{\mathbf{r}_{\mathbf{ij}}\cdot\dot{\mathbf{v}}_{\mathbf{j}}}{2c^2}\right] + \frac{1}{c^2}\sum_{j\neq i}\frac{\mu_j}{r_{ij}^3}[\mathbf{r}_{\mathbf{ij}}\cdot(4\mathbf{v}_{\mathbf{i}} - 3\mathbf{v}_{\mathbf{j}})]\mathbf{v}_{\mathbf{ij}} + \frac{7}{2c^2}\sum_{j\neq i}\frac{\mu_j\dot{\mathbf{v}}_{\mathbf{j}}}{r_{ij}} - \sum_{m=1}^{5}\frac{\mu_m\mathbf{r}_{\mathbf{im}}}{r_{im}^3} \tag{2-13}$$

where the last term includes the Newtonian effects of the five largest asteroids. Note that the acceleration depends on the position, velocity and acceleration of the other bodies. Observe that most of the general relativistic terms are of the form $(v/c)^2$. Also note that the right hand side includes accelerations, a phenomena that can not occur in Newtonian mechanics. This equation is included just to demonstrate the complexity of calculating precision ephemerides.

### 2.5.2 Approximate ephemerides

The simplest ephemerides are often given as a set of **Keplerian orbital elements** (Chapter 3). The next level of precision would also include simple time dependent variations in these elements. Over a century, errors in these simple models can be millions of kilometers for the outer planets and somewhat less for the terrestrial planets. Nevertheless, they are generally adequate for mission analysis studies and optical observations. Listed below are the orbital elements for Venus, Earth and Mars at J2000. The complete listing is given in [3,316].

Table 2-1. Planetary Orbital Elements

| **Planet** | **a** | **e** | **i** | $\Omega$ | $\varpi = \omega + \Omega$ | $\lambda_o$ |
|---|---|---|---|---|---|---|
| Venus | 0.72333199 | 0.00677323 | 3.39471 | 76.68069 | 131.53298 | 181.97973 |
| Earth-Moon Barycenter | 1.00000011 | 0.01671022 | 0.00005 | -11.26064 | 102.94719 | 100.46435 |
| Mars | 1.52366231 | 0.09341233 | 1.85061 | 49.57854 | 336.04084 | 355.45332 |

The orbits elements are from left to right, semi-major axis in **astronomical units** [3,696] (1 AU= 149,597,870.66 km), eccentricity, inclination to the mean ecliptic (Section 1.2.2) of J2000, longitude of the ascending node relative to the mean equinox of J2000, longitude of perihelion, and mean longitude at J2000=JED 2451545.0 (Section 1.3.4). The argument of perihelion is $\omega$ and $\lambda_o = M_o + \varpi$ where $M_o$ is the mean anomaly at J2000. The four angles are in degrees. Note that the ephemeris is for the Earth-Moon barycenter. The gravitational constant for the Sun [3,700] is $1.327124 \times 10^{11}$ $km^3/s^2$.

## 2.6 Problems

2-1. Show that the gravitational force on a particle inside a homogeneous spherical shell vanishes.

2-2. Find a mistake in equation (2-13)

## 2.7 Astronautics Toolbox

1. Write a function OE=PlanetOE(JD,PlanetNum), that returns the six orbital elements(OE) at the Julian date=JD. Use PlanetNum=1..9 to identify the planet and include the centennial rates. Return kilometers and radians. Get the elements from the table of Mean Orbital Elements at http://ssd.jpl.nasa.gov/

## 2.8 References


1. Moulton, F.R. **An Introduction to Celestial Mechanics**, The MacMillan Company, New York, 1914. Also available from Dover Press.
2. Estabrook, F. B. 'Post-Newtonian *n*-body Equations of the Brans-Dicke Theory,' *Astrophysics Journal*, Vol. 158, 1969.

3. Seidelmann, K. **Explanatory Supplement to the Astronomical Almanac**, University Science Books, Mill Valley, CA, 1992.

# Chapter 3 - Two Body Problem

## 3.1 Introduction

The relative motion of two particles under their mutual gravitational attraction is the corner stone of the planetary ephemerides, lunar motion, the motions of planetary moons, and artificial satellite theories. Almost all interpretations of the effects of other forces, such as non-spherical gravity fields (Section 5.4.2), N-body gravitational attraction (Section 2.3), atmospheric drag (Section 5.5.1), and solar pressure (Section 5.4.5), are described in terms of perturbations, i.e. small or slowly varying changes to the two body solution [Chapter 5].

## 3.2 Kepler's Laws

Using the relatively precise measurements of his mentor, Tycho Brahe, the essentials of two body motion were determined empirically by Kepler and captured in the three simple laws:

1. **Elliptic motion law**: The heliocentric orbit of each planet is in a fixed plane and elliptical with the Sun at one focus (1609).
2. **Equal area law**: The line from the sun to the planet sweeps out equal area in equal time (1609).
3. **Orbital period law**: The square of a planetary period is proportional to the cube of the mean distance from the Sun (1619).

Kepler tried for a number of years to fit variations of moving circles and ovals to the observations of Mars, at that time the only planet with an observable eccentric orbit. It was on the verge of quitting that he tried an ellipse with the Sun at a focus [1,141].

## 3.3 Integrals of the Two Body Problem

The equations of motion for two particles are given by equations (2-10) with n=2

$$m_1\ddot{\mathbf{r}}_1 = \frac{-Gm_1m_2(\mathbf{r}_1 - \mathbf{r}_2)}{r^3} \qquad m_2\ddot{\mathbf{r}}_2 = \frac{-Gm_1m_2(\mathbf{r}_2 - \mathbf{r}_1)}{r^3}$$

where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ defines the relative position of $m_1$ with respect to $m_2$. With $M=m_1+m_2$, the equation of relative motion of $m_1$ with respect to $m_2$ is obtained by forming $\ddot{\mathbf{r}}$ and substituting from the equations above to obtain the **fundamental equation of motion** for the two body problem

$$\ddot{\mathbf{r}} = \frac{-GM\mathbf{r}}{r^3} \tag{3-1}$$

In the two body problem GM is often represented by μ. It is noted that relative motion depends only on the total mass of the two bodies. Also the equation is symmetric, that is, the equation of motion is independent of the reference body. For many problems one of the masses is much greater than the other and there is a tendency to forget that the more massive body is also orbiting the less massive body. Equations (3-1) are a set of three second order, coupled, non-linear, homogeneous, autonomous, ordinary differential equations. The solution therefore requires six independent integrals or constants. Since the equations are non-linear, a closed form solution in terms of elementary functions is not expected. Even though the autonomous nature means the reference epoch for time in not required, in orbital mechanics, as discussed in Chapter 1, it is prudent to think of time in terms of year, month, day, hour, minute and seconds of ephemeris or dynamic time (1.3.2).

### 3.3.1 Angular momentum.

From Section 2.4.2, the total system angular momentum is conserved. The **specific relative angular momentum** (simply referred to as angular momentum) $\mathbf{h} = \mathbf{r} \times \mathbf{v}$ is also conserved. To test for conservation of angular momentum, it is natural to form the cross product of $\mathbf{r}$ with equation (3-1) to obtain $\mathbf{r} \times \dot{\mathbf{v}} = \dot{\mathbf{h}} = 0$. Clearly **angular momentum is conserved for any central force system.** Since $\mathbf{h}$ is a vector, it represents three constants of integration and one immediate implication is that the relative position and velocity vectors must lie in the plane normal to $\mathbf{h}$ and through the center of mass of the reference body. This plane is called the **orbit plane**. Both bodies move in the same plane which contains the barycenter of the system. Kepler's observation that the planetary motion is planar is thus a result of the conservation of angular momentum. The second of Kepler's laws is also derived from this result as follows. Let $\theta$ be the angular position in the orbital plane measured from an arbitrary reference line. The magnitude of the angular momentum is the radial distance times the angular component of the velocity, i.e. $h = r^2\frac{d\theta}{dt}$. But, from elementary calculus the area sweep out in time dt is $\frac{1}{2}r^2 d\theta$. Thus, **conservation of angular momentum implies that the orbital motion will sweep out equal area in equal time**. This is a verification to Kepler's equal area law (Section 3.2.). Angular momentum can also be written as $h = rv\cos\gamma$ where $\gamma$ is the **flight path angle** or angle between the velocity and the local horizontal.

### 3.3.2 Energy.

Total system energy is conserved as seen in Section 2.4.3. The test for relative **conservation of mechanical energy** is to form the rate at which the system forces are doing work i.e. $\dot{W} = \mathbf{v} \cdot \mathbf{F}$. Forming the dot product of the velocity with equation (3-1) yields

$$\mathbf{v} \cdot \dot{\mathbf{v}} = \frac{1}{2}\frac{d}{dt}(\mathbf{v} \cdot \mathbf{v}) = \frac{-\mu(\mathbf{r} \cdot \mathbf{v})}{r^3} = \mu\frac{d}{dt}\left(\frac{1}{r}\right)$$

since $r\dot{r} = \mathbf{r} \cdot \mathbf{v}$ (B-2). This leads immediately to the energy integral

$$\frac{v^2}{2} - \frac{\mu}{r} = E \qquad (3\text{-}2)$$

thus providing the fourth of six necessary constants of integration. From the form of the equation it is clear that the energy can be positive or negative because the potential energy has been referenced to $r = \infty$. If the initial conditions have a "low" velocity and a "small" radius the energy will be negative. Specifically, if initially $rv^2 < 2\mu$ then the energy is negative. In this case, the energy integral alone limits the motion to the bounded, spherical region $r < \mu/(-E)$. This spherical surface is called a **zero velocity surface** because it is a surface that can not be crossed once the energy is known. If, on the other hand, $E \geq 0$ there are no spatial limit to the motion provided by the energy integral. When E>0 equation (3-2) is often written as $v^2 = v_\infty^2 + \frac{2\mu}{r} = v_\infty^2 + v_e^2$, to show that there is a non-zero velocity as the particle approaches an infinite distance. The **velocity at infinity** is $v_\infty$ which is zero if E=0 corresponding to an infinite radius for the surface of zero velocity. The **escape velocity** or **parabolic velocity** is $v_e$, which is the minimum velocity at distance r that will provide "escape" from the central body.

Exercise 3-1. Use the energy integral to show that if the initial conditions are such that $r_o v_o^2 = \mu$, then the maximum distance between the bodies is $2r_o$.

### 3.3.3 In-plane orbit geometry

Equation (3-1) describes the three dimensional motion; but, from above it is known that the motion is in the plane normal to **h**. If **h**=**0** then **r** and **v** are co-linear and the motion is a straight line toward or away from the center of attraction. This case will be considered later. Otherwise, it is desirable to have a form of the EOM that only describes the motion in the plane. To this end, cross **h**(≠**0**) with equation (3-1) to get

$$\dot{\mathbf{v}} \times \mathbf{h} = \frac{-\mu}{r^3}(\mathbf{r} \times \mathbf{h})$$

Recalling that **h** and **r** are orthogonal and using the magnitude of **h** from above gives

$$\frac{d}{dt}(\mathbf{v} \times \mathbf{h}) = \frac{\mu}{r^3} r(r^2\dot{\theta}\mathbf{e}_\theta) = \mu\dot{\theta}\mathbf{e}_\theta = \mu\frac{d}{dt}\mathbf{e}_r$$

since $\frac{d\mathbf{e}_r}{dt} = \dot{\theta}\mathbf{e}_\theta$ (B-2). Straight forward integration yields

$$\mathbf{v} \times \mathbf{h} = \mu(\mathbf{e_r} + \mathbf{c}) \tag{3-3}$$

where **c** is the vector constant of integration. It is seen that **c** lies in the orbit plane, is dimensionless and was obtained by the integration of a vector equation. Since **c** is in the orbital plane, it does not provide three new constants of integration. As shown below **c** provides the direction of the line of apsides of the conic orbit and thus **c** only provides the fifth of the necessary integrals. This can be seen by reducing equation (3-3) to a scalar equation by forming the dot product with **r** to get

$$\mathbf{r} \cdot (\mathbf{v} \times \mathbf{h}) = \mathbf{h} \cdot \mathbf{h} = h^2 = \mu\mathbf{r} \cdot (\mathbf{e_r} + \mathbf{c}) = \mu r(1 + c\cos f)$$

where f is the angle between **r** and **c**. Solving for r gives the equation of a conic section with origin at a focus

$$r = \frac{p}{1 + e\cos f} \tag{3-4}$$

where the **semi-latus rectum** is $p=h^2/\mu$, the **eccentricity** $e = c \geq 0$ and the **true anomaly** f is the angle from the line defined by the minimum r (**periapsis**) to the current value of **r**. Note that p is completely determined by the angular momentum. For an ellipse $p=a(1-e^2)$ where a is the length of the **semi-major axis** of the ellipse. Equation (3-4) is called the **equation of the orbit** and is the mathematical statement of Kepler's elliptic motion law (Section 3.2).

Exercise 3-2. Starting with the equation of an ellipse with origin at the center, $\xi^2/a^2 + \eta^2/b^2 = 1$, show that equation (3-4) is the equation of the ellipse with origin at a focus and that $p=b^2/a$ where b is the length of the semi-minor axis.

If $p \neq 0$, the type of conic section is determined by the eccentricity. The minimum radius occurs when the denominator of equation (3-4) is a maximum i.e. when f=0. When e=0 the radius is a constant so the motion is in a circle. If e<1 there is a maximum radius at $f=\pi$, and the motion is **elliptical**. If $p \neq 0$ and e=1 the motion is **parabolic** and the radius is infinite at $f=\pi$. Finally, if e>1 the motion is **hyperbolic** and the asymptotes correspond to values of true anomaly ($f_\infty$) that make the denominator zero, where $f_\infty = \cos^{-1}(-1/e)$. There is thus a range of forbidden values of true anomaly for hyperbolic motion. For hyperbolas, the geometric definition of p is $a(e^2-1)$ with a>0. In orbital mechanics it is convenient to set $p=a(1-e^2)$ and let a<0 for hyperbolic motion. It is generally clear by inspection which convention is being used. However, computers do not have such reasoning capability, so care must be exercised in computer programs to pick a convention and use it throughout all procedures.

Figure 3-1 shows the orbit geometry for the elliptical case. The periapsis distance is $r_p=a(1-e)$, apoapsis distance $r_a=a(1+e)$, semi-latus rectum $p=a(1-e^2)$ and semi-minor axis is $b = a\sqrt{1-e^2}$. The distance from the center to the focus is ae.

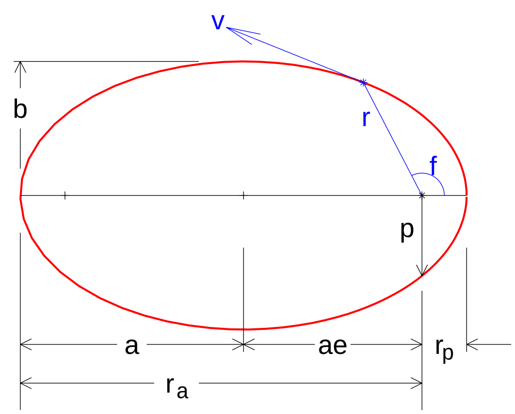



Figure 3-1. Elliptical orbit geometry

Regardless of the type of conic, at periapsis the velocity must be normal to the radius so $h^2=r_p^2 v_p^2=\mu p$. From above, $r_p=a(1-e)$ and $p=a(1-e^2)$ for elliptic and hyperbolic motion, so

$$v_p^2 = \frac{\mu}{a}\left(\frac{1+e}{1-e}\right) \qquad r_p > 0$$

Evaluating the energy integral equation (3-2) at periapsis leads to $E= -\mu/2a$. So that a more traditional version is the **vis-viva integral**

$$v^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right) \tag{3-5}$$

Exercise 3-3. Fill in the steps from equation (3-2) to equation (3-5).

By comparing equation (3-3) and the vis-viva integral at periapsis, it can be seen that $\mathbf{c}=e\mathbf{e_r}$, i.e. **c** points toward periapsis and has magnitude equal to the eccentricity. So **c** is redesignated as **e**, the **eccentricity vector** and is given by

$$\mathbf{e} = \frac{\mathbf{v} \times \mathbf{h}}{\mu} - \mathbf{e_r} \tag{3-6}$$

Note that the eccentricity vector is NOT a unit vector. It is well defined for all cases with $e \neq 0$. Since **r** can never be zero, $\mathbf{e}_r$ is always defined. If $\mathbf{h} = \mathbf{r} \times \mathbf{v} = \mathbf{0}$ the conic is **degenerate** and the motion is **rectilinear**, i.e. a straight line either toward or away from the center of attraction. In either case $\mathbf{e} = -\mathbf{e}_r$, and as would be expected the periapsis is in the opposite direction of the position vector.

Exercise 3-4. Derive equation (3-6) starting with equation (3-3).

Exercise 3-5. Draw and annotate a sketch like Figure 3-1 for the parabolic and hyperbolic cases. Show the asymptotes for the latter case.

### 3.3.4 Orbital plane orientation

The orientation of the orbit plane in three dimensional space and the location of the line of apsides in the orbit plane are usually defined by the [3,1,3] Euler rotation angles [Ω, i, ω]. These angles are illustrated in Figure 3-2 and can be calculated from the angular momentum vector and eccentricity vector. Like all three parameter representations there is a singularity, i.e. a situation is which the angles are not unique. With latitude and longitude the singularity is at the pole where longitude is undefined. For the [3,1,3] rotation, the singularity is when i=0 or π and neither Ω nor ω are uniquely defined. However, the **longitude of periapsis**, $\varpi=\Omega+\omega$ may still be well defined if $e \neq 0$ even though it is not an angle in the usual sense.

If $h \neq 0$, the orbital **inclination** is given by

$$\cos i = \mathbf{e_h} \cdot \mathbf{e_z} = \frac{h_z}{h} \qquad 0 \le i \le \pi \tag{3-7}$$

where the h=0 case is discussed in Section 3.11. An orbit with i<π/2 is said to be a **direct orbit.** An orbit with i=π/2 is called a **polar orbit** and if i>π/2 the orbit is said to be a **retrograde orbit**.

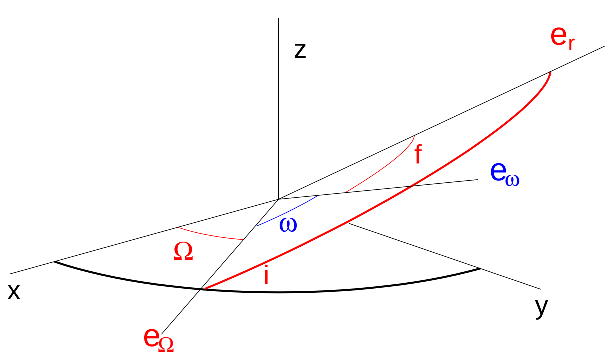


Figure 3-2. Orbit orientation elements.

The **ascending node** is the point where the particle passes through the x-y plane with positive $\dot{z}$. The unit vector in this direction is in both the x-y plane and the orbit plane and is given by

$$\mathbf{e}_\Omega = \frac{\mathbf{e_z} \times \mathbf{e_h}}{|\mathbf{e_z} \times \mathbf{e_h}|} \tag{3-8}$$

The cases where this definition does not provide a unique vector (i.e. **h**=**0** or **h**=h$\mathbf{e_z}$) are discussed in Section 3.11. The **longitude of the ascending node** is given by

$$\sin\Omega = \mathbf{e_x} \times \mathbf{e}_\Omega \cdot \mathbf{e_z} = \frac{h_x}{\sqrt{h_x^2 + h_y^2}} \qquad \cos\Omega = \mathbf{e_x} \cdot \mathbf{e}_\Omega = \frac{-h_y}{\sqrt{h_x^2 + h_y^2}} \qquad 0 \le \Omega < 2\pi \tag{3-9}$$

Which can also be written as

$$h_x = h \sin i \sin\Omega \qquad h_y = -h \sin i \cos\Omega \qquad h_z = h \cos i \tag{3-10}$$

and provide the tradition means of determining both inclination and longitude of the node.
If $e \ne 0$ define a unit vector toward periapsis $\mathbf{e}_\omega = \mathbf{e}/e$. The **argument of periapsis** is then given by

$$\sin\omega = \mathbf{e}_\omega \cdot (\mathbf{e_h} \times \mathbf{e}_\Omega) \qquad \cos\omega = \mathbf{e}_\omega \cdot \mathbf{e}_\Omega \qquad 0 \le \omega < 2\pi \tag{3-11}$$

Equations (3-5) through (3-11) can be evaluated from the initial conditions $\mathbf{r}(t_o)$ and $\mathbf{v}(t_o)$ to determine the five Keplerian elements a, e, i, Ω, and ω as five constants of integration. The sixth orbital element, and last integration constant, is developed in the next section.

### 3.3.5 Motion in the orbital plane

None of the above five integrals of the motion explicitly involve time, i.e. given a time there is no relation above that will provide the position and velocity of the body. One approach to developing a relation between time and position in orbit can be derived from the vis-viva integral and conservation of angular momentum in the form

$$v^2 = \dot{r}^2 + r^2\dot{f}^2 = \dot{r}^2 + \frac{h^2}{r^2} = \mu\left(\frac{2}{r} - \frac{1}{a}\right)$$

giving as the differential equation for r

$$\frac{dr}{dt} = \pm\frac{1}{r}\sqrt{\frac{-\mu r^2}{a} + 2\mu r - h^2} \tag{3-12}$$

This integral involves the square root of a quadratic polynomial and can therefore be integrated in terms of elementary functions to yield r as a function of time. The form of the solution depends on the sign of the coefficient of $r^2$, yielding regular trigonometric functions if the coefficient is negative (**elliptical** motion, a>0), hyperbolic functions if the sign is positive (**hyperbolic** motion, a<0), and simple functions if the coefficient is zero (**parabolic** motion, 1/a=0)

In elementary calculus one method for integrating similar forms is to replace the dependent variable by a trigonometric variable. Without justification, new variables E and F are introduced for the elliptic and hyperbolic cases:

$$\begin{array}{llll} r = a(1 - e\cos E) & a > 0 & e \le 1 & 0 \le E < 2\pi \\ r = a(1 - e\cosh F) & a < 0 & e \ge 1 & -\infty < F < \infty \end{array} \tag{3-13}$$

The elliptic **eccentric anomaly**, E, is seen to be a well defined variable permitting r to vary from a(1-e) to a(1+e) as required by the equation of the orbit, equation (3-4). Also it is seen that if $0<E<\pi$ then $0<f<\pi$ and likewise if $\pi<E<2\pi$ then $\pi<f<2\pi$. Similarly for the hyperbolic eccentric anomaly F.

From this point, only the elliptical case will be developed in detail. Since E is well defined, the definition above can be differentiate with respect to time to yield $\dot{r} = ae\sin E\dot{E}$. After this expression and the definition are substituted into equation (3-12) a little algebra leads to $(1 - e\cos E)\dot{E} = \sqrt{\mu/a^3}$. This equation can be immediately integrated to yield **Kepler's equation** for the elliptical case

$$M = n(t - \tau) = E - e\sin E \tag{3-14}$$

which provides the sixth and final constant of integration $\tau$, the **time of periapsis passage**. The **mean motion** is denoted $n = \sqrt{\mu/a^3}$. The time from one periapsis passage to the next is the **period**, P. Since E would change by $2\pi$ during this time, $P = \frac{2\pi}{n} = 2\pi\sqrt{a^3/\mu}$, which is Kepler's orbital period law (Section 3.2). In practical orbital analysis it is not uncommon for $t-\tau$ to be larger than the orbital period. Thus the analyst must be prepared for $|E| > 2\pi$. The **mean anomaly** is defined by $M=n(t-\tau)$ and describes an angle that evolves linearly with time. The mean anomaly permeates orbital mechanics but is purely for notational convenience as a surrogate for time. From Kepler's equation, the difference between mean anomaly and eccentric anomaly is periodic and is never greater than the eccentricity.

Exercise 3-6. Make the substitution (3-13) in (3-12) to verify equation (3-14).

Following the same steps for the hyperbolic case and recalling that $a<0$ and $e>1$, Kepler's equation for hyperbolic motion can be shown to be

$$M = n(t - \tau) = e\sinh F - F \tag{3-15}$$

where $n = \sqrt{-\mu/a^3}$ and if $t<\tau$ then $F<0$. The concept of orbital period is of course meaningless for this case, nevertheless the notation $M=n(t-\tau)$ is still used. Most text do not associate Kepler with this equation since he did not derive it. Nevertheless, to shorten terminology, both elliptic and hyperbolic forms will be referred to as Kepler's equation.

The parabolic case is easier to derive from the conservation of angular momentum in the form $r^2\frac{df}{dt} = h = \sqrt{\mu p}$. For a parabola, e=1, so the equation of the orbit can be written as

$r = \frac{p}{1+\cos f} = \frac{p}{2}\sec^2\frac{f}{2}$. Combining these two leads to the equation $\sec^4\frac{f}{2}\frac{df}{dt} = 4\sqrt{\mu/p^3}$ which can be integrated to yield **Barker's equation**

$$M = \sqrt{\frac{\mu}{p^3}}(t-\tau) = \frac{1}{6}\tan^3\frac{f}{2} + \frac{1}{2}\tan\frac{f}{2} \tag{3-16}$$

For notational continuity, M is also defined for parabolic motion, but the functional form is different than in Kepler's equation.

Exercise 3-7. Perform the elementary integration to derive equation (3-16).

In equations (3-14) through (3-16) the time of periapsis is an ephemeris time (Section 1.3.3) epoch often defined in either Julian day (Section 1.3.4) or YYMMDDHHMNSS.SS notation. Time must generally be carried to the microsecond level and is often represented by two numbers to maintain such accuracy. Typical representations are (1) modified Julian date (1.3.4) and seconds into the day, (2) year and seconds from beginning of the year, (3) YYYYMMDDHHMM and SS.SSS.... form and (4) year and day of the year.

This completes the development of the **classical Keplerian orbital elements** for the two body problem. For elliptical and hyperbolic motion a, e, i, Ω, ω and τ are utilized. For parabolic motion a and e are replaced by the single parameter p. Parabolic motion has only 5 independent parameters to define the orbit since it is known that e=1.

The trigonometric relationships between the true and eccentric anomalies can be derived directly from the equation of the orbit, equation (3-4), and (3-13), giving for f(E) and f(F)

$$\begin{aligned} \cos f &= \frac{\cos E - e}{1 - e\cos E} & \qquad \cos f &= \frac{e - \cosh F}{e\cosh F - 1} \\ \sin f &= \frac{\sqrt{1-e^2}\sin E}{1 - e\cos E} & \qquad \sin f &= \frac{\sqrt{e^2-1}\sinh F}{e\cosh F - 1} \end{aligned} \tag{3-17}$$

Exercise 3-8. Invert equations (3-17) to obtain E(f) and F(f) as given in Table 3-1

## 3.4 Orbital Elements from Initial Position and Velocity

The calculation of the classical Keplerian orbital elements (a, e, i, Ω, ω, τ) given **r** and **v** at time t is relatively straight forward using the equations above. Modern tracking accuracies require that double precision calculations be performed for most orbits. The issues in calculating the orbital elements are (1) when to assume a non-degenerate orbit is parabolic ($e \approx 1$), and any special consideration for (2) the circular orbit case ($e \approx 0$), (3) the low inclination case ($i \approx 0, \pi$) and (4) the **degenerate** conic case ($h \approx 0$). One approach to these issues is given in Section 3.11. For the rest of this section these subtleties will be ignored.

The following steps provide the method to calculate the orbital elements from position and velocity.

1. Use the vis-viva integral equation (3-5) to calculate $z = 1/a$, the reciprocal of the semi-major axis. The reciprocal is used since it is well defined even for parabolic orbits.
2. Calculate angular momentum and related variables $\mathbf{h} = \mathbf{r} \times \mathbf{v}$, h, and $p = h^2/\mu$.
3. Use equation (3-7) to determine inclination, i.
4. Use equations (3-9) or (3-10) to determine ascending node longitude, Ω.
5. Utilize equations (3-6) and (3-11) to determine argument of periapsis, ω.
6. Finally, τ, the time of periapsis is calculated using either equation (3-14), (3-15) or (3-16) depending on the sign of z. Quadrants are determined using r and $r\dot{r} = \mathbf{r} \cdot \mathbf{v}$ along with either:

   a. $\sin E = \dfrac{z^2 r\dot{r}}{ne}$ $\qquad \cos E = \dfrac{1}{e}(1 - zr)$ for elliptical (z>0) motion.

   b. $\sin f = \dot{r}\sqrt{\dfrac{p}{\mu}}$ $\qquad \cos f = \dfrac{p}{r} - 1$ for parabolic (z=0) motion.

   c. $\sinh F = \dfrac{z^2 r\dot{r}}{ne}$ $\qquad \cosh F = \dfrac{1}{e}(1 - zr)$ for hyperbolic (z<0) motion.

Return z, p, i, Ω, ω and τ as the element set.

Exercise 3-9. From equations in Section 3.3, develop the expression for sinE in part a. and sinf in part b.

Numerous other sets of six elements have been developed. Some of these are combinations of Kepler elements utilized to eliminate a singularity for a particular problem. For examples, P=e*sin*ω and Q=e*cos*ω have been used for low eccentricity orbits, while R=*sin*i*sin*Ω and S=*sin*i*cos*Ω have been used for low inclination orbits. Various sets of “universal variables” have also been utilized. These are valid for all three types of orbits at the cost of introducing new functions defined in terms of infinite series [2,168] or in terms of continued fraction [1,187] or simply branching and using the equations above. However, the classical elements provide physical insight into orbit geometry and are adequate with careful handling of degenerate or nearly-degenerate cases as discussed in Section 3.11.

## 3.5 Solution of Kepler's and Barker's Equations

If the position and velocity are given at some time t, then either Kepler's or Barker's equation can be used to calculate the time of periapsis. On the other hand, these equations are transcendental functions of the anomalies. So if the orbital elements are given and position and velocity at time t are desired, solution methods for Kepler’s and Barker’s equations must be developed.

Barker's equation (3-16) for parabolic motion is a cubic in tan(f/2), so a closed form solution exist from elementary algebra. Battin [1,151] gives a computationally robust solution repeated here.

Referring to (3-16) let $B = 3\sqrt{\frac{\mu}{p^3}}(t-\tau) = 3M$ $A = \left(|B| + \sqrt{1+B^2}\right)^{2/3}$ then the solution to Barker's equation is

$$\tan\frac{f}{2} = \frac{2AB}{1+A+A^2} \tag{3-18}$$

Exercise 3-10. Verify that equation (3-18) is a solution by substitution into (3-16).

For the elliptical motion case there are two popular approaches to solving Kepler's equation. The first is successive substitution

$$E_{k+1} = M + e\sin E_k \tag{3-19}$$

and the second is Newton iteration

$$E_{k+1} = E_k + \frac{M - E_k + e\sin E_k}{1 - e\cos E_k} \tag{3-20}$$

Danby [2,149] and Meeus [3,181] provide excellent discussions of iteration methods, evaluations of various starting values, and the advantages of including higher order Taylor series terms in the Newton iteration. Colwell [4] provides a history of solving this equation. Successive substitution is easiest to implement but can require 10 or more iterations even for e<0.1 and may not converge for e>0.8. The traditional starting value is $E_1 = M$, but Newton's method can become unstable due to the denominator being small when $|M|<\pi/6$ and $0.95<e<1$. However, with the proper starting condition Newton's method will converge in less than five iteration for all M and any e<1 [3,181]. Danby [2,152] suggest an initial guess of $E_1 = M+0.85\ e\ \text{sign}(\sin(M))$ which will converge in six or less Newton iterations to eleven decimal places for $0 \le e < 1$.

Exercise 3-11. Implement both equation (3-19) and equation (3-20) with different starting conditions and evaluate the convergence properties for $0<M<2\pi$ and $0.05<e<0.95$. Write a 2-3 page paper on the results. Use the starting values above and consider equation (3-23).

For the **hyperbolic** case e>1 so that successive substitution must take the form

$$F_{n+1} = \sinh^{-1}\left(\frac{n(t-\tau)+F_n}{e}\right)$$

to assure convergence. Recall that inverse hyperbolic functions can be written in terms of logarithms, for example, $\sinh^{-1}x = \log\left(x+\sqrt{1+x^2}\right)$.

## 3.6 Position and Velocity from Orbital Elements

If the orbital elements are known, then the position and velocity at any time t can be found by the following process.

a.) Based on z=1/a determine the type of orbit. Calculate M and if elliptical make $-\pi < M \le \pi$

b.) Determine the eccentricity from p and z.

c.) Solve Barker's or Kepler's equation for the anomalies, $-\pi < f, E < \pi$ or F as appropriate.

d.) Calculate r, $\dot{r}$ and $r\dot{\theta}$

e.) Determine the position vector using the [3,1,3] rotation [Ω, i, ω+f] starting with (r,0,0)

f.) Determine the velocity vector using the [3,1,3] rotation [Ω, i, ω+f] starting with $(\dot{r}, r\dot{\theta}, 0)$

The explicit transformations are

$$\mathbf{r} = r\mathbf{e_r} = r\begin{bmatrix}\cos\theta\cos\Omega - \sin\theta\sin\Omega\cos i\\ \cos\theta\sin\Omega + \sin\theta\cos\Omega\cos i\\ \sin\theta\sin i\end{bmatrix} \tag{3-21}$$

$$\dot{\mathbf{r}} = \dot{r}\mathbf{e_r} + r\dot{\theta}\mathbf{e}_\theta = \dot{r}\mathbf{e}_r + r\dot{\theta}\begin{bmatrix}-\sin\theta\cos\Omega - \cos\theta\sin\Omega\cos i\\ -\sin\theta\sin\Omega + \cos\theta\cos\Omega\cos i\\ \cos\theta\sin i\end{bmatrix} \tag{3-22}$$

where θ=ω+f corresponds to the third rotation. Note that $\mathbf{e}_\theta = \dfrac{\partial\mathbf{e}_r}{\partial\theta}$.

## 3.7 Expansions for Elliptic Motion

In the era of analytic solutions it was often necessary to make approximations to arrive at any solutions at all. Taylor series expansions are a familiar tool. For periodic orbital motion the Fourier series representations are generally more useful and there are numerous such representations in the two body problem. These are developed in detail in a number of reference books [1,206], [5,33], [6, Chapter II] and will not be developed here. Such expansions are useful for making initial estimates for iterative solutions, for obtaining approximate solutions, for making order of magnitude estimates, and in orbit perturbation problems (Section 5.4.1). A few of these expansions are given below to terms through eccentricity cubed. The inversion of Kepler's equation yields

$$E = M + 2\sum_{k=1}^{\infty} J_k(ke)\frac{\sin kM}{k}$$

where $J_k$ are Bessel functions of the first kind of order k. Bessel invented these functions for the two body problem rather than as the solution to a differential equation. Explicitly, through terms of order $e^3$

$$E = M + \left[e - \frac{e^3}{8}\right]\sin M + \left[\frac{e^2}{2}\right]\sin 2M + \left[\frac{3e^3}{8}\right]\sin 3M + O(e^4) \tag{3-23}$$

In this equation and those below, **any term in square brackets [] is a truncated infinite series**. Within the radius of convergence, the complete expansion could be used to solve Kepler's equation without iteration. But, this is impractical since iteration is faster. Nevertheless, the first few terms can be used to obtain a first estimate for the iterative solution. It is to be noted that this series does not converge rapidly for large eccentricity. A similar expansion for r

$$\frac{r}{a} = 1 + \frac{e^2}{2} - \left[e - \frac{3e^3}{8}\right]\cos M - \left[\frac{e^2}{2}\right]\cos 2M - \left[\frac{3e^3}{8}\right]\cos 3M + O(e^4) \tag{3-24}$$

can be used to directly estimate r(t) without solving Kepler's equation. Since integration of M over $2\pi$ is the same as integrating over an orbital period, note that the mean value of r over an orbit in <u>not</u> a. The expansion for true anomaly is also a double infinite sum [1,212]

$$f = M + \left[2e - \frac{e^3}{4}\right]\sin M + \left[\frac{5e^2}{4}\right]\sin 2M + \left[\frac{13e^3}{12}\right]\sin 3M + O(e^4) \tag{3-25}$$

If $e < 0.01$, as is common for LEO and many other satellites, the last two equations can be used to calculate the position and velocity to six significant figures without solving Kepler's equation. Two additional series will be used in Chapter 5.

$$\frac{r}{a}\cos f = -\frac{3}{2}e + \left[1 - \frac{3}{8}e^2\right]\cos M + \left[\frac{e}{2}\right]\cos 2M + \left[\frac{3}{8}e^2\right]\cos 3M + O(e^3) \tag{3-26}$$

$$\frac{r}{a}\sin f = \left[1 - \frac{5}{8}e^2\right]\sin M + \left[\frac{e}{2}\right]\sin 2M + \left[\frac{3}{8}e^2\right]\sin 3M + O(e^3) \tag{3-27}$$

## 3.8 F and G Functions

The solution of equation (3-1) can be written as a Taylor series expanded about some time $t_o$ with initial conditions $\mathbf{r}_o$ and $\mathbf{v}_o$ i.e. $\mathbf{r}(t) = \mathbf{r}(t_o) + \dot{\mathbf{r}}(t_o)(t - t_o) + \frac{1}{2}\ddot{\mathbf{r}}(t_o)(t - t_o) + \ldots$. Second and higher order derivatives can be eliminated using (3-1). Following Danby [2,163], let $\sigma = \mu / r_o^3$ and $\varepsilon = \dot{r}_o / r_o$, then $\ddot{\mathbf{r}}_o = -\sigma\mathbf{r}_o$ and it can be shown that $\left.\frac{d^3\mathbf{r}}{dt^3}\right|_o = 3\sigma\varepsilon\mathbf{r}_o - \sigma\mathbf{v}_o$ etc. So the series can be written [2,437] in terms of $\mathbf{r}_o$ and $\mathbf{v}_o$ and constants $\sigma$, $\varepsilon$ and $\delta = (v_o/r_o)^2$.

$$\mathbf{r}(t) = \left[1 - \frac{\sigma}{2}(t - t_o)^2 + \frac{\sigma\varepsilon}{2}(t - t_o)^3 + \ldots\right]\mathbf{r}_o + \left[(t - t_o) - \frac{\sigma}{6}(t - t_o)^3 + \ldots\right]\mathbf{v}_o$$

An expansion of this form might have been expected since two body motion takes place in the plane (or line) defined by the initial position and velocity vectors. These vectors can be therefore used as basis vectors for any motion in the plane (or line). The two series are called the F and G functions and they are analytic in some neighborhood of $t_o$ since the series has a non-zero radius of convergence. Thus the position and velocity can be written as

$$\mathbf{r}(t) = F(t,t_o)\mathbf{r}_o + G(t,t_o)\mathbf{v}_o \qquad \mathbf{v}(t) = \dot{F}(t,t_o)\mathbf{r}_o + \dot{G}(t,t_o)\mathbf{v}_o \tag{3-28}$$

Since $\mathbf{r}(t) \times \mathbf{v}(t) = \mathbf{r}(t_o) \times \mathbf{v}(t_o) = \mathbf{h}$, $F\dot{G} - G\dot{F} = 1$ and also note that these equations are valid component wise, i.e. $x(t)=F(t,t_o)x_o + G(t,t_o)\dot{x}_o$. Because of slow convergence, this form has had limited utility except as a basis for analytic approximations over short times. A more useful form can be obtained by first introducing the **orbital coordinate system** ($\xi$, $\eta$, $\zeta$). The origin of this system is the center of attraction and the fundamental plane is the orbit plane, i.e. the $\zeta$ axis is along **h.** The $\xi$ axis points to periapsis to define the fundamental direction and the $\eta$ axis completes the right hand system pointing in the direction of the velocity at periapsis. The following relations can be developed from the above

$$\begin{aligned} \xi &= r\cos f = a(\cos E - e) & \frac{d\xi}{dt} &= \frac{-na\sin f}{\sqrt{1-e^2}} = \frac{-na^2\sin E}{r} \\ \eta &= r\sin f = a\sqrt{1-e^2}\sin E & \frac{d\eta}{dt} &= \frac{na(\cos f + e)}{\sqrt{1-e^2}} = \frac{na^2\sqrt{1-e^2}\cos E}{r} \end{aligned} \tag{3-29}$$

Equations (3-28) are independent of the particular coordinate system chosen and so are equally applicable to the orbital system, i.e.

$$\xi_t = F(t, t_o)\xi_o + G(t, t_o)\dot{\xi}_o \qquad \eta_t = F(t, t_o)\eta_o + G(t, t_o)\dot{\eta}_o \tag{3-30}$$

If $\eta_t$ and $\xi_t$ are considered as being known, then these equations can be thought of as two equations in the two unknowns F and G so that

$$F(t,t_o) = \frac{1}{h}[\xi_t\dot{\eta}_o - \eta_t\dot{\xi}_o] \qquad G(t,t_o) = \frac{1}{h}[\eta_t\xi_o - \xi_t\eta_o] \tag{3-31}$$

where it is noted that the determinant of the coefficients of F and G is the angular momentum, h>0. Similar arguments can be made for the velocities. The states in the orbital coordinate system can be eliminated in favor of either the true or eccentric anomaly using equations (3-29) to yield

$$\begin{aligned} F(t,t_o) &= 1 - \frac{r}{p}[1 - \cos(f - f_o)] = 1 + \frac{a}{r_o}[\cos(E - E_o) - 1] \\ G(t,t_o) &= \frac{rr_o}{h}\sin(f - f_o) = (t - t_o) + \frac{1}{n}[\sin(E - E_o) - (E - E_o)] \\ \dot{F}(t,t_o) &= \frac{-h}{p^2}[\sin(f - f_o) + e(\sin f - \sin f_o)] = \frac{-na^2}{rr_o}\sin(E - E_o) \\ \dot{G}(t,t_o) &= 1 + \frac{r_o}{p}[\cos(f - f_o) - 1] = 1 + \frac{a}{r}[\cos(E - E_o) - 1] \end{aligned} \tag{3-32}$$

Thus the F and G functions can be determined for any times t and $t_o$ by solving Kepler's equation at the respective times to obtain E and $E_o$. The orbit position can then be propagated from any time $t_o$ to any time t using (3-28). The three Euler angles ($\Omega$, $\omega$, i) are not required in this approach.

Exercise 3-12. Utilize equations (3-29) through (3-31) to derive the first line of equations (3-32).

## 3.9 Coordinate System Rotation

The (3,1,3) rotation matrix $\Phi$ from the orbital coordinate system $\rho = (\xi, \eta, \varsigma)$ to the $\mathbf{r}$=(x,y,z) system ($\mathbf{r} = \Phi\rho$), where

$$\Phi = \begin{bmatrix} \cos\Omega\cos\omega - \sin\Omega\sin\omega\cos i & -\cos\Omega\sin\omega - \sin\Omega\cos\omega\cos i & \sin\Omega\sin i \\ \sin\Omega\cos\omega + \cos\Omega\sin\omega\cos i & -\sin\Omega\sin\omega + \cos\Omega\cos\omega\cos i & -\cos\Omega\sin i \\ \sin\omega\sin i & \cos\omega\sin i & \cos i \end{bmatrix} \quad (3\text{-}33)$$

can be determined directly using the spherical trigonometry relations given in Section 1.2.1 or from the multiplication of the three rotation matrices (B-1).

## 3.10 State Propagation

Mapping or propagating the state at time $t_o$ to some other time t is one of the most common problems in orbital mechanics. For two body motion, two common approaches are

1. Transform the state at time $t_o$ to orbital elements at $t_o$ and then transform the orbital elements to the state at time t. This process would use the X2ORB and ORB2X procedures developed for the toolbox. This approach will also permit inclusion of secular and long period variations in the orbital elements due to perturbations to be discussed in Chapter 5.
2. Determine only a, e, $\tau$ and $E_o$ from the state at time $t_o$. Solve Kepler's equation at time t for E, evaluate F, G, $\dot{F}$ and $\dot{G}$, then use equations (3-28) or the parabolic or hyperbolic equivalents to determine the mapped position and velocity. Unless orbital elements are specifically desired or orbital perturbations must be included, this approach is the preferred method and utilizes the X2X procedure developed for the toolbox.

## 3.11 Degenerate, Circular and Nearly Parabolic Orbits

Numerical calculations will generally not exactly satisfy the conditions for determining the orbital elements for degenerate, low inclination, zero eccentricity, or parabolic orbits. When the condition is “nearly” satisfied, the analyst may elect to force the condition to be satisfied exactly. For example, for nearly parabolic motion, forcing parabolic motion has the advantage that Barker's equation is easier to solve than either form of Kepler's equation when the eccentricity is nearly unity. The near circular orbit case means that the argument of periapsis $\omega$ will be poorly defined from equation (3-11) because the eccentricity vector will be the difference of two much larger vectors in (3-6). One can accept the values from (3-11) or by convention define a value to $\omega$. The zero inclination case has a similar problem in that the line of nodes is poorly defined by (3-9) because $h_x$ and $h_y$ are nearly zero. Finally, degenerate conics, i.e. h=0 can occur for elliptical, parabolic and hyperbolic orbits. As seen from (3-6), all degenerate orbits have unit eccentricity and the eccentricity vector is in the opposite direction of the position vector. For degenerate orbits, (3-4) is not valid and true anomaly is undefined. However equations (3-13) through (3-15) are still valid. Barker's equation must be derived for the degenerate parabolic motion (Problem 3-1). The following steps provide general direction for the calculation of the orbital elements in these cases. The tolerance parameter “tol” is analyst supplied and depends on the accuracy requirements of the problem and the computer. For double precision 1e-8< tol < 1e-10 might be considered.

- If |zr| < tol the semi-major axis is very large compared to the initial position so set z=0 to assure parabolic motion.
- If e < tol, put periapsis at the initial position i.e. set $\mathbf{e} = \mathbf{e}_r$ and $\tau$=time of the initial conditions.

If p/r < tol, set $\mathbf{e} = -\mathbf{e}_r$, $f = -\pi$, and $p = 0$. For parabolic motion, use the results of Problem 3-1 to determine $\tau$; otherwise, use E or F calculated from equation (3-13). Use $\dot{r}$ to remove ambiguities. There are a number of options for ascending node and inclination for rectilinear orbits. One option is to set $\mathbf{e}_\Omega = \mathbf{e}_\mathbf{x}$ and select $\mathbf{e}_h$ to assure the orbit plane passes through $\mathbf{e}_r$. Another option is to set $i=\pi/2$ and $\tan\Omega=y/x$.

- If 1-|cos(i)| < tol, set i=0 or $\pi$ and $\mathbf{e}_\Omega = \mathbf{e}_\mathbf{x}$ or $\mathbf{e}_\Omega = \mathbf{e}_r$.

## 3.12 Table of Relationships

Some of the following relations are not valid if p=0. For hyperbolic orbits a<0.

Table 3-1. Two-body Problem Relationships

| Variable | Ellipse a>0, e<1 | Parabola z=0, e=1 | Hyperbola a<0, b<0, e>1 |
|---|---|---|---|
| $v^2$ | $\mu\left(\frac{2}{r}-\frac{1}{a}\right)$ | $\frac{2\mu}{r}$ | $\mu\left(\frac{2}{r}-\frac{1}{a}\right) = \frac{2\mu}{r}+v_\infty^2$ |
| $r$ | $\frac{p}{1+e\cos f} = a(1-e\cos E)$ | $\frac{p}{1+\cos f} = \frac{p}{2}\sec^2\frac{f}{2}$ | $\frac{p}{1+e\cos f} = a(1-e\cosh F)$ |
| $n$ | $\sqrt{\frac{\mu}{a^3}}$ | $\sqrt{\frac{\mu}{p^3}}$ | $\sqrt{\frac{-\mu}{a^3}}$ |
| $M = n(t-\tau)$ | $E - e\sin E$ | $\frac{1}{2}\tan\frac{f}{2}+\frac{1}{6}\tan^3\frac{f}{2}$ | $e\sinh F - F$ |
| cos f,sin f | $\frac{\cos E - e}{1-e\cos E}, \frac{\sqrt{1-e^2}\sin E}{1-e\cos E}$ | cos f, sin f | $\frac{e-\cosh F}{e\cosh F-1}, \frac{\sqrt{e^2-1}\sinh F}{e\cosh F-1}$ |
| cosE, coshF | $\frac{e+\cos f}{1+e\cos f}$ | NA | $\frac{e+\cos f}{1+e\cos f}$ |
| sinE, sinhF | $\frac{\sqrt{1-e^2}\sin f}{1+e\cos f}$ | NA | $\frac{\sqrt{e^2-1}\sin f}{1+e\cos f}$ |
| $\frac{dE}{dt}, \frac{dF}{dt}$ | $\frac{na}{r}$ | NA | $-\frac{na}{r}$ |
| $\frac{dr}{dt}$ | $\frac{na^2 e\sin f}{b} = \frac{na^2 e\sin E}{r}$ | $np\sin f$ | $-\frac{na^2 e\sin f}{b} = \frac{na^2 e\sinh F}{r}$ |
| $\xi$ | $r\cos f = a(\cos E - e)$ | $r\cos f$ | $r\cos f = a(\cosh F - e)$ |
| $\eta$ | $r\sin f = b\sin E$ | $r\sin f$ | $r\sin f = -b\sinh F$ |
| $\frac{d\xi}{dt}$ | $\frac{-na^2\sin f}{b} = \frac{-na^2\sin E}{r}$ | $-np\sin f$ | $\frac{na^2\sin f}{b} = -\frac{na^2\sinh F}{r}$ |
| $\frac{d\eta}{dt}$ | $\frac{na^2(\cos f+e)}{b} = \frac{nab\cos E}{r}$ | $\frac{np^2}{r}$ | $\frac{-na^2(\cos f+e)}{b} = \frac{nab\cosh F}{r}$ |
| $F(t,t_0)$ | $\frac{1}{h}(\xi_t\dot{\eta}_o - \eta_t\dot{\xi}_o)$ where $\xi_t = \xi(t)$, etc. | | |
| $G(t,t_0)$ | $\frac{1}{h}(\eta_t\xi_o - \xi_t\eta_o)$ | | |
| $\dot{F}(t,t_0)$ | $\frac{1}{h}(\dot{\xi}_t\dot{\eta}_o - \dot{\eta}_t\dot{\xi}_o)$ | | |
| $\dot{G}(t,t_0)$ | $\frac{1}{h}(\dot{\eta}_t\xi_o - \dot{\xi}_t\eta_o)$ | | |

## 3.13 Problems

3-1. Starting with equation (3-12), derive a form of Barker's equation (3-16) for degenerate parabolic motion.

3-2. Show that for small $|t-t_o|$, equation (3-32) reduces to the expected limit.

3-3. Develop the equivalent of equations (3-32) for parabolic orbits.

3-4. Develop the equivalent of equations (3-32) for hyperbolic orbits.

3-5. Verify the $\Phi(2,2)$ term in equation (3-33) using spherical trigonometry relations.

## 3.14 Astronautics Toolbox

1. Write a procedure that returns the rotation matrix (3 by 3) for an arbitrary [3,1,3] set of rotations $[\alpha,\beta,\gamma]$, $\Phi$=Rotate313($\alpha,\beta,\gamma$,ichk).
2. Write a procedure to solve Barker's equation (3-16), f=Barker(t,$\tau$,p,$\mu$,ichk). Assume t is (n by 1).
3. Write a procedure to solve Kepler's equation (3-14) for elliptic motion using Newton-Raphson iteration, E=Kepler(M,e,tol,ichk). Assume M is (n by 1) and "tol" is the relative error in E for convergence.
4. Write a procedure to solve Kepler's equation (3-15) for hyperbolic motion using Newton-Raphson iteration, F=KeplerH(M,e,tol,ichk). Assume M is (n by 1) and "tol" is the relative error in F for convergence.
5. Write a procedure to transform from rectangular coordinates to orbital elements for any type of motion. [OE]=X2Orb(t,**r**,**v**,$\mu$,ichk), where **r** and **v** are given at a single time t and OE is the six vector $(z,p,i,\Omega,\omega,\tau)$.
6. Write a procedure to provide position and velocity at an array of times for any type of orbit. [**r**,**v**]=Orb2X(t,OE,$\mu$,tol,ichk) where t is (n by 1), OE is the six elements used above, and "tol" is the relative accuracy for convergence of Kepler's equation. Output position and velocity are both (n by 3).
7. Write a procedure, using the F and G function approach, to transform from an initial state at time $t_1$ to states at an array of times t. [**r**,**v**]=X2X(t,$t_1$,$\mathbf{r_1}$,$\mathbf{v_1}$,$\mu$,tol,ichk), where, t (n by 1) and **r** and **v** are (n by 3).

## 3.15 References


1. Battin, R.H., **An Introduction to the Mathematics and Methods of Astronautics**, AIAA Education Series, 1987.
2. Danby, J.M.A., **Fundamentals of Celestial Mechanics**, Willmann-Bell, Richmond, Va., 1989.
3. Meeus, John, **Astronomical Algorithms**, Willmann-Bell, Inc., Richmond, Va., 1991.

4. Colwell, Peter, **Solving Kepler's Equation over Three Centuries**, Willmann-Bell, Inc., Richmond, Va., 1993.
5. Plummer, H.C., **An Introductory Treatise on Dynamical Astronomy**, Constable and Company Limited, 1918. Also available from Dover, 1960.
6. Brouwer, D. and Clemence, G., **Methods of Celestial Mechanics**, Academic Press, 1961

# Chapter 4 - Three Body Problem

## 4.1 Introduction

The three body problem has been of considerable interest for centuries because the Earth-Moon-Sun system can be approximated as a three body problem. During the Apollo era, the problem received renewed interest since the two dominate forces action on the spacecraft were due to the Earth and the Moon. Even for this system, requiring 18 integrals, a closed-form solution of the **general problem** does not appear feasible and there are no known integrals beyond those discussed in Chapter 2. The three body problem was recognized by Poincare as being what we now call a chaotic system, i.e. the characteristics of the motion are very sensitive to the initial conditions. There exist, however, particular solutions of the three-body problem obtained by Lagrange in 1772, which will be discussed later.

## 4.2 Restricted Problem

Of interest in the problem of three bodies is the special case in which the mass m of one of the bodies is so small that its motion does not affect the motion of the two remaining bodies. The motion of the two massive bodies $m_1$ and $m_2$ is obtained from the solution of the two-body problem and can be assumed to be known. The problem is further restricted by considering the case in which $m_1$ and $m_2$ move in circular orbits about their barycenter with constant angular velocity $\omega$ whereas the infinitesimal mass m moves under the combined gravitational attraction of both $m_1$ and $m_2$. Under these circumstances, the problem is reduced to the investigation of a 3-degree of freedom (DOF) system. This problem is called the **restricted problem of three bodies**. If m is further restricted to the plane of motion of $m_1$ and $m_2$, there is a 2-DOF system.

The classical coordinate system has the origin at the barycenter and the fundamental plane is the plane of motion of the two finite bodies. From equations (2-10) the equations of motion are

$$\ddot{\rho} = -Gm_1\frac{\rho - \boldsymbol{\rho_1}}{\rho_{01}^3} - Gm_2\frac{\rho - \boldsymbol{\rho_2}}{\rho_{02}^3} \tag{4-1}$$

where $\rho_{0i}$ is the distance from m to $m_i$ and $\rho_i$ is the position vector of $m_i$. Select as the unit of length the constant distance between $m_1$ and $m_2$ and the unit of mass so that $m_1+m_2=1$. It is readily shown (Section 3.3.5) that the mean motion of the two finite masses is unity, i.e. $n=\omega=1$.

Now transform to a rotating coordinate system so that the two finite masses remain on the x axis. Let **r** be the position vector in the rotating system, then the transformation is

$$\rho = \begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{r}$$

Using this expression to transform the above equations of motion to the rotating frame yields

$$\ddot{x} - 2\dot{y} = x - (1-\mu)\frac{x - x_1}{r_1^3} - \mu\frac{x - x_2}{r_2^3}$$
$$\ddot{y} + 2\dot{x} = y - (1-\mu)\frac{y}{r_1^3} - \mu\frac{y}{r_2^3} \tag{4-2}$$
$$\ddot{z} = -(1-\mu)\frac{z}{r_1^3} - \mu\frac{z}{r_2^3}$$

where wolog $\mu \le 0.5$ is the normalized mass of $m_2$, $x_1$=-μ is the location of $m_1$ and $x_2$=1-μ is the location of $m_2$ on the x axis, and $r_i$ is the distance from m to $m_i$. These equations can be written as

$$\ddot{x} - 2\dot{y} = \frac{\partial U}{\partial x} \qquad \ddot{y} + 2\dot{x} = \frac{\partial U}{\partial y} \qquad \ddot{z} = \frac{\partial U}{\partial z} \tag{4-3}$$

where the pseudo-force function is defined by

$$U = \frac{x^2 + y^2}{2} + \frac{1-\mu}{r_1} + \frac{\mu}{r_2} \tag{4-4}$$

The latter two terms come from the gravity potential and the first term comes from the "**centrifugal potential**."

Exercise 4-1. Fill in the steps from equation (4-1) to equation (4-2) for the x-component

Exercise 4-2. Verify that equations (4-3) and (4-4) are equivalent to equation (4-2)

The only integral of this system is an energy type integral discovered by Jacobi. To seek an energy integral, multiply each of equations (4-2) by the corresponding velocity component and add the three equations. The sum is integrable and leads to **Jacobi's integral**

$$\dot{x}^2 + \dot{y}^2 + \dot{z}^2 = v^2 = 2U - C = x^2 + y^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} - C \tag{4-5}$$

The constant C is called **Jacobi's constant**. Although this is the only known integral of the six required, it has proven very useful in studying orbital motion in systems that can be approximated by the restricted problem.

Exercise 4-3. Fill in the steps to develop equation (4-5) starting with equation (4-2).

#### 4.2.1 Jacobi's integral and Tisserand's criteria

It is believed that the Oart cloud is the source of observed comets. This cloud is well outside the orbit of Pluto. Comets in the cloud are generally in nearly circular orbits about the Sun. However, close encounter between two comets or perturbations from nearby stars or Jupiter could significantly reduce the angular momentum of a comet so that the new perihelion is a few astronomical units or less. Many years later the comet may pass sufficiently close to the Sun and the Earth that the ablating ice is visible from the Earth. The comet will either be in a periodic orbit about the Sun or the encounter in the Oart cloud added enough energy that the comet will subsequently escape the solar system. Observed periodic comets will therefore be in orbits with eccentricities near unity and large semi-major axes. Most of the orbital period is spent in the outer part of the solar system, otherwise the comet would have long ago been destroyed by the Sun's radiant energy.

If the elliptical orbit of a comet is unperturbed by one of the planets, subsequent appearances of the comet can be identified by the two body orbital elements about the Sun. On the other hand, if the comet is perturbed by a single planet then Jacobi's integral can be applied to the Sun, the perturbing planet, and the comet three body system so that the comet can be identified by Jacobi's constant. Tisserand found a simple way of relating Jacobi's constant to the Keplerian heliocentric orbital elements. Transform Jacobi's integral back to the non-rotating $(\xi,\eta,\zeta)$ system to get

$$\left(\frac{d\xi}{dt}\right)^2 + \left(\frac{d\eta}{dt}\right)^2 + \left(\frac{d\zeta}{dt}\right)^2 - 2\left(\xi\frac{d\eta}{dt} - \eta\frac{d\xi}{dt}\right) = \frac{2(1-\mu)}{\rho_{01}} + \frac{2\mu}{\rho_{02}} - C \tag{4-6}$$

where $\rho_{0i}$ is defined above. For planets $\mu<<1$ so that the first three terms can be interpreted as the velocity relative to the Sun and the next two terms as the $\zeta$ component of angular momentum relative to the orbital plane of the planet about the Sun.

Exercise 4-4. Fill in the steps to develop equation (4-6) starting with equation (4-5)

Using the vis-viva integral equation (3-5) and $h^2 = a(1-e^2)$ reduces Jacobi's integral to

$$\frac{2}{\rho_{01}} - \frac{1}{a} - 2\sqrt{a(1-e^2)}\cos i = \frac{2}{\rho_{01}} + \frac{2\mu}{\rho_{02}} - C$$

where 1-$\mu$ has been set to unity on both sides of the equation. If Jacobi's constant is evaluated when the comet is far from the perturbing planet $\left(\frac{2\mu}{\rho_{02}} \ll \frac{1}{a}\right)$ this equation becomes

$$\frac{1}{a} + 2\sqrt{a(1-e^2)}\cos i = C$$

Evaluating this relation before and after the encounter with the perturbing planet yields **Tisserand's criteria**

$$\frac{1}{a_1} + 2\sqrt{a_1(1-e_1^2)}\cos i_1 = \frac{1}{a_2} + 2\sqrt{a_2(1-e_2^2)}\cos i_2 \tag{4-7}$$

for the identification of comets that have been perturbed by a single planetary encounter. Based on the before and after heliocentric orbits, the planet at which the close encounter occurred can generally be identified. If not, since the semi-major axis and inclination in the equation are calculated according to the encountered planet, a number of planets must be tested before the identification of the comet can be confirmed or rejected based on this criteria.

### 4.2.2 Zero velocity surfaces

The value of C in equation (4-5) can be determined from a set of initial conditions. If C>0 equation (4-5) places a constraint on the possible spatial locations of the trajectory. In particular, motion can only occur in regions where $v^2 \geq 0$. Recall that equation (3-2) limits the possible spatial locations for the two body problem. For a given C the surface defined by $v^2=0$ is called the **zero velocity surface**. Motion can only occur on one 'side' of the zero velocity surface.

From analytic geometry, a single equation relating the three spatial coordinates x, y, and z defines a two dimension subspace. For example, $x^2+y^2=R^2$ defines the surface of an infinite circular cylinder of radius R with the z-axis along the center of the cylinder, while $x^2+y^2<R^2$ defines the three dimensional space inside the cylinder. Similarly, $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2} = 1$ defines the surface of an ellipsoid with principal axes a, b, and c along x, y and z.

Moulton [1,281] provides methods for calculating and an extensive discussion of the zero velocity surfaces. Figure 4-1 shows some of the zero velocity contours in the plane z=0 for μ=0.25 and From equation (4-5) it is seen that motion can only occur in a region where

$$x^2 + y^2 + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} > C \qquad (4\text{-}8)$$

For example, if the initial conditions have $z=\dot{z}=0$ and result in C=5, then motion is confined to the x-y plane in the nearly circular region about either mass or to the region outside the C=5 outer contour. If one wanted to design a trajectory that goes from a point near $m_1$ to $m_2$, then C must be less than about 3.87. Similarly, if the 2-d motion of m is initiated near either mass and C>3.56 then m can never escape from the figure-eight region defined by the C=3.56 contour. Finally, applications of Jacobi's integral like the above are most useful for defining where motion **cannot** occur. There is no guarantee, that for a specified value of C, an actual trajectory exists between every two points in the region bounded by the C contour.

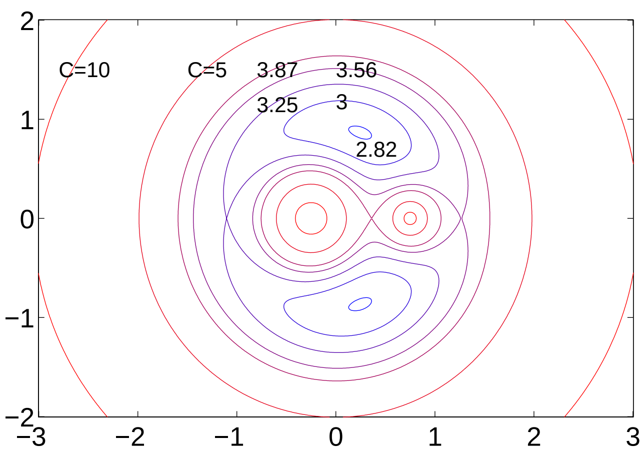


Figure 4-1. Zero velocity contours, μ=0.25.

Exercise 4-5. Use the Matlab meshdom, mesh and contour functions to generate zero velocity surfaces and contour plots in the x-z plane for μ=0.25. Same scale as Figure 4-1. Interpret results.
Exercise 4-6.

### 4.2.3 Lagrange points

Lagrange discovered that there were five equilibrium points for the restricted three body problem. These positions correspond to solutions (x, y, z) to equations (4-2) when the velocity and acceleration terms are zero. In the inertial system, the particle at such an equilibrium point, would be in a circular orbit about the center of mass of $m_1$ and $m_2$. At an **equilibrium point,** equations (4-2) can be written as

$$x-(1-\mu)\frac{x-x_1}{r_1^3}-\mu\frac{x-x_2}{r_2^3} = -x\left(\frac{1-\mu}{r_1^3}+\frac{\mu}{r_2^3}-1\right)-(1-\mu)\mu\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right) = 0$$

$$y\left(\frac{1-\mu}{r_1^3}+\frac{\mu}{r_2^3}-1\right) = 0 \tag{4-9}$$

$$z\left(\frac{1-\mu}{r_1^3}+\frac{\mu}{r_2^3}\right) = 0$$

The first equation is factored in two ways for later use. The third of these equations implies that all equilibria must be in the x-y plane. The second equation admits to two types of solutions, y=0 and $r_1$=$r_2$=1. Solutions with y=0 must therefore be on the x-axis and satisfy the condition

$$f(x) = x-(1-\mu)\frac{x-x_1}{|x-x_1|^3}-\mu\frac{x-x_2}{|x-x_2|^3} = 0 \tag{4-10}$$

The zeros or roots of f(x) are the equilibrium points. Note that f(x)>0 for sufficiently large and positive x. As x becomes smaller and approaches $x_2$, the gravity potential term for $m_2$ dominates and $f(x_2^+)<0$. Thus there is exactly one equilibrium point with $x>x_2=1-\mu$ denoted $L_2$. When x is slightly less than $x_2$ this potential term still dominates but is now positive so $f`(x_2^-)>0$. But as x becomes smaller and approaches $m_1$, this potential dominates so that $f(x_2^+)<0$. Thus there is exactly one equilibrium point between the two masses called $L_1$. By the same arguments, there is exactly one equilibrium point with $x<x_1=-\mu$ called $L_3$. These three equilibria are called the **straight line solutions** since all three masses remain in a line. Moulton [1] provides power series expansions in $\sqrt[3]{\mu}$ for calculating the locations along the x axis for L1 and L2. Retaining only the first term in the series, the distance from $m_2$ to L1 and L2 is $\tilde{\rho}_2 = \sqrt[3]{\mu/3}$. Likewise, if $\mu \ll 1$, $\tilde{\rho}_1 = 1-\tilde{\rho}_2$ is the distance from $m_1$ to L3. Newton-Raphson iteration works effectively for finding the roots of equation (4-10) if μ has a numerical value.

Exercise 4-7. Make a single Matlab plot of f(x’) from equation (4-10) for μ=[0.1:0.1:0.5] and x’ = [-2:0.02:2], where the origin for x’ is at mass 1-μ and mass μ is at x’=1. Note this is not the same coordinate system use in equation (4-10). Interpret results.

The second set of solutions to the y-equation, i.e. $r_1=r_2=1$ can be easily shown to also satisfy the x-equation for equilibrium. These two points, $\left(\frac{1}{2}-\mu, \mp\frac{\sqrt{3}}{2}, 0\right)$, are called the **equilateral triangle solutions** and denoted $L_4$ and $L_5$. In Figure 4-1, $L_1$ is where the C=3.87 contour crosses itself between the masses, $L_2$ is where the C=3.56 contour crosses near x=1.2, and $L_3$ is located where the C=3.25 contour crosses near x=-1. $L_4$ and $L_5$ are inside the C=2.82 contour.

The Lagrange solutions of the restricted three-body problem are of more than purely academic interest. If the Sun-Earth system is considered, satellites have been located at $L_1$ [2] to permit measurements of the solar wind before it arrives at the Earth and produces changes in the ionosphere and the geomagnetic field. The ionosphere is important for low frequency radio transmission and over-the-horizon radar. Disruptions in the ionosphere can be very dramatic during solar storms. Further, the electrical power distribution system, on numerous occasions, has had major black outs over large geographical areas when the geomagnetic field has changed drastically during a solar storm. One astrophysical phenomena which has been attributed to these solutions is the **Gegenschein** (counterglow). The Gegenschein is a faint glow observed at night in a position exactly opposite the sun and may result from reflection of sunlight off dust that is near the Earth-Sun equilibrium position $L_3$.

For the Sun and Jupiter system, there are a number of asteroids, called the **Trojan asteroids**, oscillating about $L_4$ or $L_5$. For the Earth-Moon system there have been numerous studies of placing a relay satellite near $L_2$ but sufficiently far away that the satellite could be seen from the Earth. Though not at the equilibrium point, the unbalanced forces acting on the satellite would be small and this position would require limited station keeping propulsion. There are “**halo orbits**” about these equilibria that have been exploited for various scientific purposes [2]. There was also a report in the early 1960's that clouds of dust were observed near $L_4$. This dust was attributed to a contemporary meteor impact on the back side of the Moon. Though these observation were never independently confirmed, numerous simulations were performed to study the possibility.

### 4.2.4 Stability of Lagrange points

After determining the existence of equilibrium points, the next issue is to determine the stability of each point. Only linearized stability will be considered here so that no conclusions will be drawn about global stability. Let $(x_o, y_o, z_o)$ be an equilibrium point and $(\xi, \eta, \zeta)$ be small deviations from equilibrium, i.e. $x=\xi+x_o$, $y=\eta+y_o$ and $z=\zeta+z_o=\zeta$. Linearizing equations (4-3) about the equilibrium point yields

$$\begin{aligned} \ddot{\xi}-2\dot{\eta} &= U_{xx}\xi+U_{xy}\eta+U_{xz}\zeta \\ \ddot{\eta}+2\dot{\xi} &= U_{yx}\xi+U_{yy}\eta+U_{yz}\zeta \\ \ddot{\zeta} &= U_{zx}\xi+U_{zy}\eta+U_{zz}\zeta \end{aligned} \tag{4-11}$$

where U is given by equation (4-4), the usual notation for second derivatives is used, and it is understood that the derivatives are evaluated at the equilibrium point under study. Equations (4-

11) are a set of coupled, second order, autonomous, homogeneous, ordinary differential equations so the solution is a sum of exponentials. The **characteristic equation**, which completely defines the dynamics of this system, is obtained by substituting $\xi(t)=\xi_o e^{\lambda t}$, $\eta(t)=\eta_o e^{\lambda t}$, $\zeta(t)=\zeta_o e^{\lambda t}$ into equation (4-11) and seeking non-trivial solutions. Before performing this operation, note that U has continuous derivatives (except at two uninteresting points) and is symmetric in z, therefore $U_{zx}$ and $U_{zy}$ vanish in the x-y plane. In this case, equations (4-11) shows that the perturbed motion in the z direction is uncoupled from the motion in the x-y plane and reduces to

$$\ddot{\zeta} - U_{zz}\big|_{z=0}\zeta = \ddot{\zeta} + \omega_z^2\zeta = 0 \tag{4-12}$$

where $\omega_z^2 = \frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3} > 0$. At all five equilibrium points, motion in the z direction is uncoupled and harmonic. Also note that at $L_4$ and $L_5$ $\omega_z=1$, which is the same as the mean motion of $m_1$ and $m_2$, so that z-only linearized motion produces a closed orbit in 3-d inertial space as well as in the rotating system. In the inertial system, the motion describes a nearly circular orbit with a small inclination to the x-y plane.

Exercise 4-8. Starting with equations (4-3) develop the $\xi$ component of equations (4-11)

The characteristic equation for perturbed motion in the x-y plane is obtained from the first two of equations (4-11) and reduces to the bi-quadratic

$$\lambda^4 + (4 - U_{xx} - U_{yy})\lambda^2 + U_{xx}U_{yy} - U_{xy}^2 = 0 \tag{4-13}$$

(4-14)

where all of the coefficients are evaluated at the particular equilibrium point under study. For stable solutions, $\lambda$ must be pure imaginary so that $\lambda^2$ must be negative. Let $\nu=\lambda^2$ and write equation (4-13) as

$$\nu^2 + b\nu + c = 0$$

From the quadratic formula it is clear that the roots will be negative only if $b>0$, $c>0$ and $b^2>4c$. For any of the straight line solutions $y=z=0$ and by symmetry $U_{xy}=U_{yz}=U_{xz}=0$. Unfortunately, symmetry arguments are not applicable to the two remaining terms and analysis must be performed to show that $U_{xx} = 1+ 2\omega_z^2$ and $U_{yy} = 1- \omega_z^2$, so $b = 2 - \omega_z^2$ and $c = (1+2\omega_z^2)(1-\omega_z^2)$. The sign of c is determined from the sign of $1- \omega_z^2$, which from the second form of the first of equations (4-9) can be written as

$$1 - \omega_z^2 = \frac{(1-\mu)\mu}{x}\left(\frac{1}{r_1^3} - \frac{1}{r_2^3}\right)$$

For $L_1$ and $L_2$, x>0 and both points are closer to $m_2$ than $m_1$ so c will be negative at these two points. At $L_3$ x<0 and $L_3$ is closer to $m_1$ than $m_2$ so c is also negative at this point. Thus all of the straight line solutions are unstable.

Exercise 4-9. Starting with equations (4-11) develop equation (4-13)

At $L_4$ and $L_5$, $\omega_z$=1 and the x-y characteristic equation (4-13) reduces to

$$\lambda^4 + \lambda^2 + \frac{27\mu(1-\mu)}{4} = \nu^2 + \nu + \frac{27\mu(1-\mu)}{4} = 0$$

By Descartes rule of signs there are no positive roots and either 0 or 2 negative roots for ν. So if there are real roots they must be negative. Using the notation above, the condition for stable motion is $b^2$>4c which reduces to

$$\mu < \frac{1}{2} - \sqrt{\frac{23}{108}} = 0.03852\ldots \approx \frac{1}{25.96}$$

Thus, the triangle equilibrium points are stable if the primary to secondary mass ratio is greater than about 24.96. All Sun/planet and planet/moon pairs in the solar system satisfy this criteria except for Pluto/Claron. Finally, it must be remembered that any particular Lagrange point may not be stable when other forces are included in the equations of motion or the motion of the primaries is not circular.

## 4.3 Finite Mass Particular Solutions

A number of particular solutions for the case of three finite masses have been found. These are generalizations of the straight line and triangle solution of Lagrange discussed above. Motion of all three bodies occurs in the same plane but the distance between all three of the bodies can vary with time. From equations (2-10) the equations of motion for three bodies are

$$\ddot{\mathbf{r}}_{\mathbf{i}} = -\sum_{j \neq i}^{3} \frac{\mu_j}{r_{ij}^3}\mathbf{r}_{\mathbf{ij}} \qquad i = 1, 2, 3 \tag{4-15}$$

where $\mu_i$=G$m_i$. Wolog let the origin be at the barycenter so that

$$\sum_{i=1}^{3} \mu_i \mathbf{r}_{\mathbf{i}} = 0$$

### 4.3.1 Equilateral triangle solution

The generalization of the $L_4$ and $L_5$ equilateral triangle solution is to seek solutions where the three finite mass bodies form the vertices of an equilateral triangle at all times. However, the length of a side of the triangle is not necessarily constant. To determine if such a solution exist set $\rho=r_{ij}$ to be the equal distance between the bodies as shown in Figure 4-2. Further let $\mu=\mu_1+\mu_2+\mu_3$. Then the center of mass relation can be written as $\mu\mathbf{r_1} = \mu_2(\mathbf{r_1} - \mathbf{r_2}) + \mu_3(\mathbf{r_1} - \mathbf{r_3})$. Using this expression in equation (4-15) and to determine $r_1(\rho)$ yields the equation of motion for $m_1$

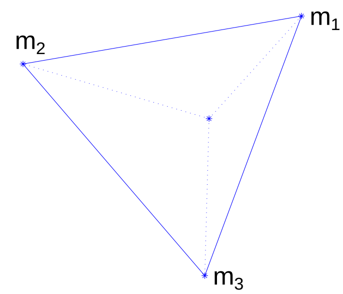



Figure 4-2. Equilateral triangle solution configuration.

$$\ddot{\mathbf{r}}_1 = -M_1\frac{\mathbf{r_1}}{r_1^3} \tag{4-16}$$

where $M_1 = \frac{(\mu_2^2 + \mu_2\mu_3 + \mu_3^2)^{3/2}}{\mu^2}$ is a reduced gravitational mass. Equations of motion for the other two masses can be obtained by cyclic permutation. This is of course the equation of motion for a particle about a center with gravitational attraction $M_1$ and located at the origin. Thus the motion is a conic as studied in Section 3.3.3. The rest of the development assumes elliptical motion, but the conclusions are equally applicable for parabolic or hyperbolic motion for the three bodies.

Exercise 4-10. Verify equations (4-16)

For the three bodies to remain in an equilateral triangle configuration the initial conditions must be chosen so that the masses have the same period, i.e.

$$\frac{M_1}{a_1^3} = \frac{M_2}{a_2^3} = \frac{M_3}{a_3^3}$$

Hence, if the period of one mass is specified, the semi-major axis for each orbit can be calculated from the above. Also the angular rate must be the same so that the angles between the masses as measured at the center of mass remain constant, i.e.

$$\frac{h_1}{r_1^2} = \frac{h_2}{r_2^2} = \frac{h_3}{r_3^2}$$

where $h_i^2 = M_ip_i$. This condition requires each mass to have the same true anomaly at any specified time. Thus all masses are at the periapsis of their respective orbits at the same time. At periapsis

$$\frac{r_i^2}{h_i} = \frac{a_i^2(1-e_i)^2}{\sqrt{M_ia_i(1-e_i^2)}} = P\frac{(1-e_i)^{3/2}}{\sqrt{1+e_i}}$$

where P is the orbital period common to all masses. The multiplier of P on the right is a monotone function of eccentricity on the interval (0,1), so the eccentricity of all three orbits is the same.

In summary, the dynamics of three finite bodies in an equilateral triangle configuration is completely defined by three mass values, an orbital period, an orbital eccentricity, a time of periapsis passage, the orientation of the orbit plane, and three arguments of periapsis that differ by 120°. More analytic detail can be found in Moulton [1,313] and Danby {3,266].

### 4.3.2 Straight line solution

To investigate the conditions for planar, straight line solutions, assume wolog that the plane of motion of the three masses is the x-y plane and the barycenter is at the origin. Let $\theta$ define the location of the line of centers in the plane, $r_i$ define the location of each mass along the line of centers relative to the barycenter, and $r_i^o$ be the location at the initial time $t_o$. For the barycenter to remain fixed, variations in position along the line must keep the ratio of distances from the center of mass constant. So introduce the time dependent variable $\rho$ such that $r_i(t) = \rho(t) r_i^o$. With $\mathbf{e_r}$ being the unit vector along the line of centers, $\mathbf{r_i} = \rho r_i^o \mathbf{e_r}$. By direct differentiation the familiar expression

$$\ddot{\mathbf{r}}_\mathbf{i} = (\ddot{\rho} - \rho\dot{\theta}^2) r_i^o \mathbf{e_r} + \frac{1}{\rho}\frac{d}{dt}(\rho^2\dot{\theta}) r_i^o \mathbf{e_\theta}$$

is derived. Substituting into equations (4-15) gives for $m_1$

$$\ddot{\rho} - \rho\dot{\theta}^2 = \left[\frac{\mu_2}{r_1^o}\frac{(r_2^o - r_1^o)}{\left|r_2^o - r_1^o\right|^3} + \frac{\mu_3}{r_1^o}\frac{(r_3^o - r_1^o)}{\left|r_3^o - r_1^o\right|^3}\right]\frac{1}{\rho^2} = -\frac{M_1}{\rho^2} \tag{4-17}$$

$$\frac{d}{dt}(\rho^2\dot{\theta}) = 0$$

These are of course the equation of motion for the two body problem. Since $\rho$ is dimensionless, $M_1$ has dimensions of an angular rate squared, so denote $M_1 = \omega_1{}^2$. Angular momentum for each mass is again preserved. The equivalent mass or mean motion in the radial equation must be the same regardless of which mass is used to derive the equation of motion for $\rho$, i.e. $M_1 = M_2 = M_3$ and $\omega_i = \omega$, i=1,2,3. If three initial positions are specified consistent with the barycenter location, then the period of the motion and the initial velocity for $m_1$ can be obtained from

$$\omega^2 r_1^o - \mu_2\frac{(r_2^o - r_1^o)}{\left|r_2^o - r_1^o\right|^3} - \mu_3\frac{(r_3^o - r_1^o)}{\left|r_3^o - r_1^o\right|^3} = 0$$

and the other two initial velocities can be obtained by cyclic permutation. Multiplying each equation in turn by $\mu_i$ and adding yields $\mu_1 r_1^o + \mu_2 r_2^o + \mu_3 r_3^o = 0$, the condition for the invariance of the center of mass. Thus the three conditions are not independent and the set of necessary conditions can be written by taking the first two expressions and the center of mass requirement [1,310], i.e.

$$
\begin{aligned}
\mu_1 x_1 + \mu_2 x_2 + \mu_3 x_3 &= 0 \\
\omega^2 x_1 - \mu_2(x_1 - x_2) - \mu_3 \frac{x_1 - x_3}{r_{13}^3} &= 0 \\
\omega^2 x_2 - \mu_1(x_2 - x_1) - \mu_3 \frac{x_2 - x_3}{r_{23}^3} &= 0
\end{aligned}
\tag{4-18}
$$

where for notational simplification $x_i \equiv r_i^o$. Wolog let $x_1<x_2<x_3$ and select the unit of length such that $r_{12}=1$ then equations (4-18) reduce to

$$
\begin{aligned}
\mu_1 x_1 + \mu_2(1 + x_1) + \mu_3 x_3 &= 0 \\
\mu_2 + \frac{\mu_3}{(x_3 - x_1)^2} + \omega^2 x_1 &= 0 \\
-\mu_1 + \frac{\mu_3}{(x_3 - x_1 - 1)^2} + \omega^2(1 + x_1) &= 0
\end{aligned}
$$

Using the first equation to eliminate $x_3$ from the second and third equations and then eliminating $\omega^2$ between the remaining two equations yields

$$
\begin{aligned}
&(\mu x_1 + \mu_2 + \mu_3)^2(\mu x_1 + \mu_2)^2[\mu_2 + (\mu_1 + \mu_2)x_1] \\
&+ \mu_3^3(1 + x_1)(\mu x_1 + \mu_2 + \mu_3)^2 - \mu_3^3 x_1(\mu x_1 + \mu_2)^2 = 0
\end{aligned}
\tag{4-19}
$$

where $\mu=\mu_1+\mu_2+\mu_3$ is the total gravitational constant. A little algebra will show that this quintic equation for $x_1$ has all positive coefficients. By Descartes rule of signs there are no positive roots and one, three or five negative roots. This result does not provide much new information since it is already known that $x_1$ must be negative. However, if $x_1$ is eliminated from equation (4-19) in favor of $x_{32} \equiv x_3 - x_2$ using

$$
\mu x_1 + \mu_2 + \mu_3 + \mu_3 x_{32} = 0
$$

then it can be shown that $x_{32}$ must satisfy

$$
\begin{aligned}
&(\mu_1 + \mu_2)x_{32}^5 + (3\mu_1 + 2\mu_2)x_{32}^4 + (3\mu_1 + \mu_2)x_{32}^3 \\
&- (\mu_2 + 3\mu_3)x_{32}^2 - (2\mu_2 + 3\mu_3)x_{32} - (\mu_2 + \mu_3) = 0
\end{aligned}
\tag{4-20}
$$

This quintic has exactly one positive root so there is only one root with $x_3>x_2$. After this root is found, $x_1$ can be determined from the above. Then $x_2=1+x_1$ and finally $x_3=x_{32}+x_2$. The common value of $\omega^2$ can be obtained by either the second or third of equations (4-18). Three solutions to equation (4-20) can be obtained by cyclic permutation; but, these are the same as simply rearranging the mass values. After the $x_i$ are found, the solution can be scaled to real dimensions. Initial velocities can be calculated from the equations above. Perhaps the simplest approach is to select $r_i^o = a_i$ so that $(1-e) \le \rho \le (1+e)$ and calculate the velocities at periapsis, i.e. $\rho$=1-e. The velocity at periapsis is given by $v_i = (\mp a_i)\omega\sqrt{\frac{1+e}{1-e}}$, where the sign is selected depending on the location of $m_i$ with respect to the center of mass.

## 4.4 Problems..

4-1. Use the Matlab meshdom, mesh and contour functions to generate zero velocity surfaces and contour plots in the (x-y), (x-z) and (y-z) planes for the Earth-Moon system. Same scale as Figure 4-1. Compare to Figure 4-1 and interpret results.

4-2. Write a Matlab procedure using ODE 45 to solve equations (4-2). Apply to motion near $L_4$ or $L_5$ in the Earth-Moon system. Provide five example plots of the trajectories in the x-y plane for five different initial conditions with $z=\dot{z}=0$. Vary initial conditions to show the transition from bound motion to unbound motion. At least one case should verify that your solution at the libration point is correct. Interpret results.

4-3. Provide a semi-log or log plot of the period of oscillation for the two solutions to equation (4-13) over the range of mass ratios from 0.001 to 0.5. Interpret results.

## 4.5 Astrodynamics Toolbox

1. Write a function x=XLn($\mu$,n,tol,ichk) that returns the x location of the n-th (n=1,2,3) Lagrange point for mass ratio $\mu$ to an accuracy specified by tol.
2. Write a function f=LnFreq($\mu$,n,ichk) that returns the two frequencies of oscillation at Lagrange point n (n=1:5) for mass ratio $\mu$.

## 4.6 References


1. Moulton, F.R., **An Introduction to Celestial Mechanics**, The MacMillan Company, New York, 1914. Also available from Dover
2. Howell, K.C. and Pernicka, H.J., “Station Keeping Method for Libration Point Trajectories,” *Journal of Guidance, Control and Dynamics,* Vol. 16, No. 1, January-February, 1993, pp. 151-159.
3. Danby, J.M.A., **Fundamentals of Celestial Mechanics**, Willmann-Bell, Richmond, Va., 1989.

# Chapter 5 - Orbital Perturbations

## 5.1 Introduction

The motion of planets and natural or artificial satellites can be approximated by modeling both bodies as point masses and assuming that the only forces acting between them are the mutual gravitational attraction. The relative motion is then described by the two body solution discussed in Chapter 3. There are numerous additional forces affecting the relative motion. Both the additional forces and the deviations from the two body motion are called **perturbations**. When these forces are small compared to the central gravitational attraction, they may cause only small and/or slow deviations from two body motion and might be addressed analytically. The analytic solutions can be used as computationally efficient approximations to the motion or perhaps more importantly, can provide insight into the effects of the perturbations. By adding a disturbing force onto equation (3-1), the equation for the motion of $m_1$ relative to $m_2$ can be written as

$$\ddot{\mathbf{r}} + \mu\frac{\mathbf{r}}{r^3} = \mathbf{f} \tag{5-1}$$

where the perturbing force is $\mathbf{f}$. Note the abuse of language here: $\mathbf{f}$ is actually the relative acceleration produced by whatever physical process is causing the deviation from two body motion. The relative acceleration is the acceleration produced on $m_1$ minus the acceleration produced on $m_2$. In addition to the gravitational attraction of other masses presented in Chapter 2, these perturbations can come from numerous sources and their effects on the orbit vary greatly. Perturbing forces include aerodynamic interaction with the atmosphere, electromagnetic interactions with the magnetic field and charged particle belts, and gravitational forces due to the non-spherical gravity field of the central body. Perturbations are produced by the momentum flux of electromagnetic energy from the Sun called **radiation pressure** and particle flux called the **solar wind**. Reflected and radiated flux from the Earth can also produce significant perturbations on low Earth orbit (LEO) satellites.

When a **numerical** solution is sought to (5-1) or an equivalent form of the EOM, the approach is called the method of **special perturbations.** Special perturbation methods will only be discussed briefly in Section 5.6. If an **analytic** solution is sought, the approach is called the method of **general perturbations**. General perturbation methods are usually based on a form of (5-1) that is derived using the variation of parameters method from the theory of ordinary differential equations.

## 5.2 Variation of Parameters

As a basis for developing Lagrange's planetary equations (5-18)**,** the **variation of parameters** method for solving differential equations is reviewed by applying the approach to a harmonic oscillator with a non-linear spring. The equation of motion is

$$\ddot{x} + \omega^2 x = -\varepsilon x^3 \qquad \varepsilon > 0 \qquad \varepsilon \ll 1 \tag{5-2}$$

Letting $\rho = x$, $\nu = \dot{x}$ gives the first order form

$$\dot{\rho} = \nu \qquad \dot{\nu} = -\omega^2\rho - \varepsilon\rho^3 \tag{5-3}$$

For $\varepsilon$=0 the solution is harmonic motion with period $2\pi/\omega$, amplitude A, and phase ø

$$\rho(t, A, \phi) = A\sin(\omega t + \phi) \qquad \nu(t, A, \phi) = A\omega\cos(\omega t + \phi) \tag{5-4}$$

where A and $\phi$ are determined from the initial conditions.

The variation of parameters method can be thought of as nothing more than a change in variables. In this case the dependent variables $\rho$ and $\nu$ are replaced by A and ø. It is clear from (5-4) that the transformation is well defined in both directions, except for the trivial solution A=0. To derive the equations of motion for the new dependent variables, (5-4) are differentiated with respect to time to yield

$$\begin{aligned} \dot{\rho} &= A\omega\cos(\omega t + \phi) + \sin(\omega t + \phi)\dot{A} + A\cos(\omega t + \phi)\dot{\phi} \\ \dot{\nu} &= -A\omega^2\sin(\omega t + \phi) + \omega\cos(\omega t + \phi)\dot{A} - A\omega\sin(\omega t + \phi)\dot{\phi} \end{aligned} \tag{5-5}$$

Using (5-4) and (5-5) to eliminate $\rho$ and $\nu$ from (5-3) yields

$$\begin{aligned} \sin(\omega t + \phi)\frac{dA}{dt} + A\cos(\omega t + \phi)\frac{d\phi}{dt} &= 0 \\ \cos(\omega t + \phi)\frac{dA}{dt} - A\sin(\omega t + \phi)\frac{d\phi}{dt} &= \frac{-\varepsilon A^3}{\omega}\sin^3(\omega t + \phi) \end{aligned} \tag{5-6}$$

Using the fact that the coefficient matrix of the derivatives is non-singular along with a few trigonometry identities, (5-6) can be written as

$$\begin{aligned} \frac{dA}{dt} &= \frac{-\varepsilon A^3}{\omega}\cos(\omega t + \phi)\sin^3(\omega t + \phi) = \frac{-\varepsilon A^3}{8\omega}[2\sin 2(\omega t + \phi) + \sin 4(\omega t + \phi)] \\ \frac{d\phi}{dt} &= \frac{\varepsilon A^2}{\omega}\sin^4(\omega t + \phi) = \frac{3\varepsilon A^2}{8\omega} - \frac{\varepsilon A^2}{8\omega}[4\cos 2(\omega t + \phi) - \cos 4(\omega t + \phi)] \end{aligned} \tag{5-7}$$

These equations are exact. That is, let A(t) and ø(t) be solutions to (5-7). When these functions are substituted into (5-4) the results will be solutions to (5-3). Since (5-7) are much more complicated than (5-3), it might be said that nothing has been gained. Certainly, implementing a numerical solution to equation (5-2) would be less error prone than implementing a numerical solution to equation (5-7). But consider the case when $\varepsilon \ll 1$. In this case, from (5-7) both A and ø will change slowly with time. Assume that over one period of oscillation A and ø change so little that they can be considered constants on the right hand side of (5-7). The equations can then be integrated to show that over one period the net change in A is $\Delta A = 0$ and the net change in ø is $\Delta ø = 3\pi A^2/4\omega^2$. So that the **first order** effects of this non-linear spring are to

1. produce only **periodic** variations in the amplitude of the oscillation and

2. produce **periodic** and **secular** variations in phase. That is, the phase continues to increase a small amount each period of oscillation and this secular drift increases with amplitude.

Note that the secular drift in phase is equivalent to an amplitude dependent frequency of oscillation.

If a new variable $\phi' = \phi - \Delta\phi$ is defined, then $\phi'$ will only have periodic variation. This is the standard approach for dividing the perturbations into secular and periodic terms and is the approach for describing the motion of the vernal equinox as precession and nutation in Section 1.2.2.

Such insights into the motion are difficult to discern from (5-3). Second and higher order effects can be obtained by using the first order solutions as approximate solutions to equation (5-7) and performing another variation of parameter procedure. The next section applies the variation of parameters approach to the perturbed two body problem.

## 5.3 Lagrange's Planetary Equations

To study the effects of n-body perturbations on the motion of a planet, Lagrange applied the method of variation of parameters to equation (2-10) in the form of equation (5-1). It is for this reason that the results are given the name **Lagrange's planetary equations**. If the perturbing force $\mathbf{f}$ is derivable from a force function R then $\mathbf{f} = \nabla R$. This is of course the case for all gravitation perturbing forces. Both forms will be carried in the development. The equation numbers in brackets {} refer to the similar equation in Section 5.2. To begin, write equations (5-1) in the first order form with the non-perturbing force derivable from a potential V {(5-3)}

$$\frac{d\mathbf{r}}{dt} = \mathbf{v} \qquad \frac{d\mathbf{v}}{dt} + \nabla V(\mathbf{r}) = f(\mathbf{r}, \mathbf{v}, t) \qquad \text{or} \qquad \nabla R(\mathbf{r}) \tag{5-8}$$

where $V=-\mu/r$ and the argument indicates the coordinate system to which the gradient operator applies.

The solution to these equations can be formally written as {(5-4)}

$$\mathbf{r}(t) = \mathbf{r}(t, \mathbf{c}) \qquad \mathbf{v}(t) = \mathbf{v}(t, \mathbf{c}) \tag{5-9}$$

where $\mathbf{c}$ is the six vector of Kepler orbital elements or any other six independent constants of integration. Because neither the position nor the velocity can be written explicitly as a function of time (recall equations (3-14), (3-15), and (3-16)) one must be careful with the explicit and implicit derivatives required below. The solution for $\mathbf{r}$ and $\mathbf{v}$ are given as implicit functions of time by equations (3-21) and (3-22) in which r and f are the terms that are functions of time.

Applying the variation of parameter approach to the first of equations (5-8) yields three equations

$$\frac{d\mathbf{r}}{dt} = \frac{\partial \mathbf{r}}{\partial t} + \frac{\partial \mathbf{r}}{\partial \mathbf{c}}\frac{d\mathbf{c}}{dt} = \mathbf{v}$$

where it is now assumed that $\mathbf{r}$ is a function of t and $\mathbf{c}$ {(5-5)}. But, $\frac{\partial \mathbf{r}}{\partial t} = \mathbf{v}$ since the explicit dependence of $\mathbf{r}$ on time satisfies the equations of motion. Likewise, the second of equations (5-8) yields three more equations

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial \mathbf{c}}\frac{d\mathbf{c}}{dt} = -\nabla V(\mathbf{r}) + \mathbf{f} \text{ or } -\nabla V(\mathbf{r}) + \nabla R(\mathbf{r})$$

Again the partial time derivative must satisfy the unperturbed equations of motion. So $\frac{\partial \mathbf{v}}{\partial t} = -\nabla V$. The six differential equations of motion for $\mathbf{c}$ are therefore {(5-6)}

$$\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\frac{d\mathbf{c}}{dt} = 0 \qquad \frac{\partial \mathbf{v}}{\partial \mathbf{c}}\frac{d\mathbf{c}}{dt} = \mathbf{f} \quad \text{or} \quad \nabla R(\mathbf{r}) \tag{5-10}$$

This form of the equations of perturbed motion are not convenient because the 3 by 6 coefficient matrices on the left are functions of time. Lagrange noted that this problem can be circumvented by multiplying the first equation by $\left[\frac{\partial \mathbf{v}}{\partial \mathbf{c}}\right]^T$ and the second by $\left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right]^T$ and then subtracting to get

$$\mathbf{L}\frac{d\mathbf{c}}{dt} = \left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right]^T \mathbf{f}(c, t) \qquad \text{or} \qquad \nabla R(\mathbf{c}) \tag{5-11}$$

where $\mathbf{L} = \left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right]^T\left[\frac{\partial \mathbf{v}}{\partial \mathbf{c}}\right] - \left[\frac{\partial \mathbf{v}}{\partial \mathbf{c}}\right]^T\left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right]$ is a 6 by 6 skew symmetric coefficient matrix, R($\mathbf{r}$) has been replaced by R($\mathbf{c}$) using equation (5-9), and the gradient operator on the right creates a six vector of partials of R wrt each component of $\mathbf{c}$.

The most important property of $\mathbf{L}$ is that it is **not** an explicit function of time, that is, it can be evaluated at any point in the orbit and the same numerical values will result. To show this independence

$$\left[\frac{\partial \mathbf{L}}{\partial t}\right] = \left[\frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial \mathbf{r}}{\partial t}\right)\right]^T\left[\frac{\partial \mathbf{v}}{\partial \mathbf{c}}\right] + \left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right]^T\left[\frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial \mathbf{v}}{\partial t}\right)\right] - \left[\frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial \mathbf{v}}{\partial t}\right)\right]^T\left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right] - \left[\frac{\partial \mathbf{v}}{\partial \mathbf{c}}\right]^T\left[\frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial \mathbf{r}}{\partial t}\right)\right]$$

where the partials with respect to t and $\mathbf{c}$ have been commuted in each term. The first and fourth terms cancel since $\frac{\partial \mathbf{r}}{\partial t} = \mathbf{v}$ from above. The second term is the transpose of the third term and substituting $\frac{\partial \mathbf{v}}{\partial t} = -\nabla V(\mathbf{r})$ into the third term yields

$$-\left[\frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial \mathbf{v}}{\partial t}\right)\right]^T\left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right] = \left[\frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\right)\right]^T\left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right] = \frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial V(\mathbf{c})}{\partial \mathbf{c}}\right)^T\left(\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right) = \frac{\partial}{\partial \mathbf{c}}\left(\frac{\partial V(\mathbf{c})}{\partial \mathbf{c}}\right)^T \tag{5-12}$$

which is a symmetric 6 by 6 matrix. Thus, when the force driving the unperturbed motion is derivable from a potential, all terms cancel and **L** is not an explicit function of time. To emphasize this result, equation (5-11) might be written as

$$\mathbf{L}(\mathbf{c})\frac{d\mathbf{c}}{dt} = \left[\frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right]^T \mathbf{f}(\mathbf{c}, t) \qquad \text{or} \qquad \nabla R(\mathbf{c}) \tag{5-13}$$

Exercise 5-1. Write explicitly every term in the first row of equation (5-13) for both the **f** and R cases.

### 5.3.1 Lagrange brackets

The individual components of **L** are called **Lagrange brackets** and denoted $[c_i, c_j]$, i.e.

$$\mathbf{L}_{ij} \equiv [c_i, c_j] = \frac{\partial \mathbf{r}}{\partial c_i}^T \frac{\partial \mathbf{v}}{\partial c_j} - \frac{\partial \mathbf{v}}{\partial c_i}^T \frac{\partial \mathbf{r}}{\partial c_j} \tag{5-14}$$

As seen from this definition, the Lagrange brackets satisfy

$$[c_i, c_i] = 0 \qquad [c_i, c_j] = -[c_j, c_i] \qquad \frac{\partial}{\partial t}[c_i, c_j] = 0$$

which means there are no more than 30 non-zero brackets and only 15 have to be evaluated. Since all types of two body motion have a periapsis, **L** is traditionally evaluated at periapsis.

### 5.3.2 Rectangular coordinates

To compute the Lagrangian brackets, an appropriate **c** must be chosen. One set that results in the simplest forms of **L** is to chose the rectangular position and velocities at some epoch as the constants of integration. Though not usually thought of as constants of the motion, it is clear that they satisfy the necessary conditions of linear independence and are certainly sufficient to determine the state at any time. Wolog, let the epoch be $t_o$ and specify the conditions at epoch as the six vector $\mathbf{c} = [x_o, y_o, \ldots \dot{z}_o]^T$. From equation (5-14) the non-zero Lagrange brackets above the diagonal of **L** are

$$[x_o, \dot{x}_o] = [y_o, \dot{y}_o] = [z_o, \dot{z}_o] = 1$$

and the remaining 12 brackets are zero. Thus $\mathbf{L} = \begin{bmatrix} \mathbf{0}_{3x3} & \mathbf{I}_{3x3} \\ -\mathbf{I}_{3x3} & \mathbf{0}_{3x3} \end{bmatrix}$ and the equations of motion (5-11) in terms of the force function reduce to the **canonical** form

$$\frac{dx_o}{dt} = -\frac{\partial R}{\partial \dot{x}_o} \qquad \frac{d\dot{x}_o}{dt} = \frac{\partial R}{\partial x_o}$$

with similar expressions for the other two coordinates. There are numerous sets of choices of **c** that will lead to the canonical form; in particular, Lagrangian generalized coordinates and conjugate momenta lead to this form.

### 5.3.3 Keplerian orbital elements

A slight modification of the classical Keplerian elements leads to **c**=(a, e, i, Ω, ω, λ), where λ= -nτ to simplify the partial derivatives. To evaluate the Lagrange brackets the position and velocity must be written in terms of these elements. Equations (3-21) and (3-22) provide the necessary relationships. These equations provide the explicit dependence of position and velocity on Ω, ω, and i. The dependence on a, e, and λ are implicit through r, f, $\dot{r}$ and $\dot{f}$. There are numerous methods for evaluating the brackets and the most extensive discussions are given in Battin[5], Fitzpatrick[3], and Moulton[7]. The development below is typical.

First equations (3-21) and (3-22) are written in the orbital coordinate system (Section 3.8) using the direction cosines from equation (3-33) and relations from Table 3-1

$$\begin{aligned} \mathbf{r} &= \Phi\rho = \Phi\left[a(\cos E - e),\ \ b\sin E,\ \ 0\right]^T \\ \mathbf{v} &= \Phi\dot{\rho} = \Phi\left[-na^2\sin E/r,\ \ nab\cos E/r,\ \ 0\right]^T \end{aligned} \tag{5-15}$$

where Φ is a function only of the orientation angles (Ω, ω, i) and the position and velocity in the orbital system (Section 3.8) are functions only of a, e, and λ. Note that some of the partials of the Φ matrix yield terms in Φ, e.g.

$$\frac{\partial\Phi_{11}}{\partial\Omega} = -\Phi_{21} \qquad \frac{\partial\Phi_{11}}{\partial\omega} = \Phi_{12}$$

It is convenient to identify the columns of Φ as vectors

$$\Phi = \left[\vec{\phi}_1\ \vec{\phi}_2\ \vec{\phi}_3\right]$$

The Lagrange brackets will be evaluated at periapsis, t = τ. Since Φ is independent of time, for any orientation angle α = Ω, ω or i

$$\left.\frac{\partial\mathbf{r}}{\partial\alpha}\right|_\tau = \frac{\partial\Phi}{\partial\alpha}\rho(\tau) \qquad \left.\frac{\partial\mathbf{v}}{\partial\alpha}\right|_\tau = \frac{\partial\Phi}{\partial\alpha}\dot{\rho}(\tau)$$

But at periapsis, $\rho(\tau) = \left[r_p, 0\ ,0\right]^T$ and $\dot{\rho}(\tau) = \left[0, nab/r_p\ ,0\right]$, so it is straight forward to show that

$$\frac{\partial \mathbf{r}}{\partial \Omega} = r_p\begin{bmatrix}-\Phi_{21} & \Phi_{11} & 0\end{bmatrix}^T \qquad \frac{\partial \mathbf{v}}{\partial \Omega} = \frac{nab}{r_p}\begin{bmatrix}-\Phi_{22} & \Phi_{12} & 0\end{bmatrix}^T$$
$$\frac{\partial \mathbf{r}}{\partial \omega} = r_p\vec{\phi}_2 \qquad \frac{\partial \mathbf{v}}{\partial \omega} = -\frac{nab}{r_p}\vec{\phi}_1 \tag{5-16}$$
$$\frac{\partial \mathbf{r}}{\partial i} = r_p \sin\omega\vec{\phi}_3 \qquad \frac{\partial \mathbf{v}}{\partial i} = \frac{nab\cos\omega}{r_p}\vec{\phi}_3$$

Exercise 5-2. Starting with equation (5-15) verify the first line of equations (5-16) and (5-17)

It remains to take the partials with respect to a, e and $\lambda$. The rotation matrix $\Phi$ is not a function of these variables. So in equations (5-15), a and e appear explicitly and a, e and $\lambda$ appear implicitly through E. The implicit relationship is defined purely by Kepler's equation (3-14)

$$nt + \lambda = E - e\sin E$$

where of course n is a function of a only. From which it is easy to show that when evaluated at periapsis

$$\frac{\partial E}{\partial a} = \frac{3\lambda}{2r_p} \qquad \frac{\partial E}{\partial e} = 0 \qquad \frac{\partial E}{\partial \lambda} = \frac{a}{r_p}$$

Combining the implicit and explicit derivatives leads to

$$\frac{\partial \mathbf{r}}{\partial a} = \frac{r_p}{a}\vec{\phi}_1 + \frac{3b\lambda}{2r_p}\vec{\phi}_2 \qquad \frac{\partial \mathbf{v}}{\partial a} = -\frac{3na^2\lambda}{2r_p^2}\vec{\phi}_1 - \frac{bn}{2r_p}\vec{\phi}_2$$
$$\frac{\partial \mathbf{r}}{\partial e} = -a\vec{\phi}_1 \qquad \frac{\partial \mathbf{v}}{\partial e} = \frac{na^3}{br_p}\vec{\phi}_2 \tag{5-17}$$
$$\frac{\partial \mathbf{r}}{\partial \lambda} = \frac{ab}{r_p}\vec{\phi}_2 \qquad \frac{\partial \mathbf{v}}{\partial \lambda} = -\frac{na^3}{r_p^3}\vec{\phi}_1$$

When the expression from equations (5-16) and (5-17) are substituted into equation (5-14) it is found that there are only 6 non-zero Lagrange brackets and these are [5,482]

$$[a, \Omega] = -\frac{nb}{2}\cos i \qquad [a, \omega] = -\frac{nb}{2} \qquad [a, \lambda] = -\frac{na}{2}$$
$$[e, \Omega] = \frac{nea^3}{b}\cos i \qquad [e, \omega] = \frac{nea^3}{b} \qquad [i, \Omega] = nab\sin i$$

Equations (5-13) in terms of the non-zero Lagrange brackets are

$$[i, \Omega]\frac{d\Omega}{dt} = \frac{\partial R}{\partial \mathbf{i}}$$

$$[\lambda, a]\frac{da}{dt} = \frac{\partial R}{\partial \lambda}$$

$$[\omega, a]\frac{da}{dt} + [\omega, e]\frac{de}{dt} = \frac{\partial R}{\partial \omega}$$

$$[e, \Omega]\frac{d\Omega}{dt} + [e, \omega]\frac{d\omega}{dt} = \frac{\partial R}{\partial e}$$

$$[a, \Omega]\frac{d\Omega}{dt} + [a, \omega]\frac{d\omega}{dt} + [a, \lambda]\frac{d\lambda}{dt} = \frac{\partial R}{\partial \mathbf{a}}$$

$$[\Omega, a]\frac{da}{dt} + [\Omega, e]\frac{de}{dt} + [\Omega, i]\frac{di}{dt} = \frac{\partial R}{\partial \Omega}$$

This is a set of linear algebraic equations with constant coefficients, so as long as the matrix of coefficients is not rank deficient, these equation can be easily solved to have only time derivatives on the left. The equations of motion for a and Ω can be found by division. Time derivatives for elements e and ω can then be obtained by elimination. Finally, the equation of motion for i and λ are also obtained by elimination to yield the **Lagrange's planetary equations** {(5-7)}[5,483]

$$\begin{aligned}
\frac{da}{dt} &= \frac{2}{na}\frac{\partial R}{\partial \lambda} \\
\frac{de}{dt} &= \frac{-b}{na^3e}\frac{\partial R}{\partial \omega} + \frac{b^2}{na^4e}\frac{\partial R}{\partial \lambda} \\
\frac{di}{dt} &= -\frac{1}{nab\sin i}\frac{\partial R}{\partial \Omega} + \frac{\cos i}{nab\sin i}\frac{\partial R}{\partial \omega} \\
\frac{d\Omega}{dt} &= \frac{1}{nab\sin i}\frac{\partial R}{\partial i} \\
\frac{d\omega}{dt} &= -\frac{\cos i}{nab\sin i}\frac{\partial R}{\partial i} + \frac{b}{na^3e}\frac{\partial R}{\partial e} \\
\frac{d\lambda}{dt} &= -\frac{2}{na}\frac{\partial R}{\partial a} - \frac{b^2}{na^4e}\frac{\partial R}{\partial e}
\end{aligned} \tag{5-18}$$

These are exact equations of motion and equivalent to equation (5-8). Even if other parameters are chosen for the orbital elements, the resulting EOM are called **Lagrange's planetary equations**. Other choices might depend on the particular orbit being analyzed. For example, to derive equations (5-18) division by e and sin(i) has been performed, hence application to orbits where either of these terms is zero or nearly zero must be done with care or the non-singular elements must be utilized. As mentioned in Section 3.4, if e is small e and ω might be replaced with P=e sinω and Q=e cosω to obtain new equations of motion, [4, 337].

## 5.4 Perturbations Derivable from a Potential

The two typical perturbations derivable from a potential function are the contributions due to other point masses as in the n-body problem (Chapter 2) and the contributions due to the gravity field of the primary being non-central. The former is discussed briefly in Section 5.5.2 and the latter is developed below.

### 5.4.1 Non-spherical gravity potential

The external gravity field of most bodies can not be represented as arising from a point mass. However sufficiently large, slowly rotating bodies will closely approximate a sphere because internal shear stresses due to self gravity cannot be supported. The external gravity field potential for any body satisfies Laplace's equation, $\nabla^2 V = 0$. For nearly spherical bodies, it is natural to represent the potential in spherical coordinates $V(r,\lambda,\phi)$. Solution by separation of variables leads to the **spherical harmonic** representation, which is used here in the form

$$V(r, \phi, \lambda) = -\frac{GM}{r}\left[1 + \sum_{n=1}^{\infty}\left(\frac{R}{r}\right)^n \sum_{m=0}^{n}(C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\sin\phi)\right] \tag{5-19}$$

where $\lambda$ is longitude, $\phi$ is geocentric latitude, and R is a reference radius usually taken as the mean equatorial radius. $P_{nm}$ is the associated Legendre polynomial of degree n and order m [1].

The terms $\cos m\lambda\ P_{nm}(\sin\phi)$ and $\sin m\lambda\ P_{nm}(\sin\phi)$ are called **surface spherical harmonics** of **degree** n and **order** m. The $C_{nm}$ and $S_{nm}$ are called the **spherical harmonic coefficients** and are the unknowns that would be selected to fit the boundary conditions to obtain the solution to Laplace's equation. The reference model in the Explanatory Supplement [2,226] is actually a force function and is the negative of equation (5-19). Some representations also change the sign between the '1' and the double sum. Thus care must be exercised by the analyst to check sign conventions. The expansion above is also “unnormalized” in that the relative importance of the terms on the orbit is not directly related to the numerical value. Various normalization approaches have been [2,226] used so that the numerical value is a direct measure of the “average” acceleration produced by the term. The complete set of surface spherical harmonics are divided into three sub classes:

1. **zonal harmonics** with m=0 are rotationally symmetric about the pole and have n zero crossings from pole to pole. Note that $S_{n0}=0$. The zonal coefficients are often represented by J's, i.e. $J_n = -C_{n0}$. Here the minus sign is used, but some authors will use a plus sign. $J_2$ is the “oblateness” and $J_3$ is the “pear shape” parameter. For planets with rotational rates sufficiently large to significantly affect the surface shape, $J_2$ is greater than zero and is the dominate perturbation. The first three values of $J_n$ for Earth are

   $J_2 = 0.108263 \times 10^{-2}$ $\quad J_3 = -0.253 \times 10^{-5}$ $\quad J_4 = -0.162 \times 10^{-5}$

   For m=0, the Legendre polynomials are written without the m=0 subscript

$P_0(x)=1$ $P_1(x)=x$ $P_2(x)=(3x^2-1)/2$

where subsequent terms can be obtained from the recursion relation

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

Note this is not a particularily useful recursion formula since errors in $P_n$ produce errors that could be twice as large in $P_{n+1}$.

2. **sectorial harmonics** with n=m have no zero crossings from pole to pole since $P_{nn} \propto \cos^n\phi$, but have 2m zeros in longitude due to the *sin*m$\lambda$ and *cos*m$\lambda$ terms.

3. **tesseral harmonics** with n≠m>0 have n-m zeros from pole to pole due to $P_{nm}$ and 2m zeros in longitude due to the *sin*m$\lambda$ and *cos*m$\lambda$ terms. Recursion relations can be used [1] to increase the order so that all tesseral harmonics can be generated once the zonal harmonics are known from the recursion equation above.

In summary, **all of the surface spherical harmonics have n-m zeros pole to pole and 2m zeros in longitude**. Chobotov, Danby and other books presents graphical representation of some of these functions. In trajectory packages the gradient of the spherical harmonic functions are required. The gradient can be related to other harmonics and are also calculated recursively [1].

The spherical harmonic coefficients can be related to the **inertia integrals** of the body [3]. An inertia integral is a generalization of the traditional moments of inertia. The general inertial integral is defined by

$$I_{pqr} = \iiint \rho(x, y, z)x^p y^q z^r dxdydz$$

where the integral is taken over the physical limits of the body and $\rho$ is density. The moment of inertia about the x-axis is $I_{xx} = I_{020}+I_{002}$. Coefficients of degree n can be written as linear combinations of inertia integrals with p+q+r=n, for example $J_2 = \dfrac{I_{zz} - \frac{1}{2}(I_{xx} + I_{yy})}{MR^2}$ where M is the total mass and R is the mean radius of the body used in equation (5-19). From this form it is easier to see that $J_2$ is positive for oblate spheroids, i.e. most planets and other large, rotating bodies. Now $J_1 = \dfrac{I_{001}}{MR} = \dfrac{\bar{z}}{R}$, where $\bar{z}$ is the z component of the center of mass location. From this result, all the first degree coefficients are zero if the origin is taken at the center of mass. In precision Earth satellite orbit determination programs, the center of mass of the Earth is permitted to deviate from the coordinate system origin to account for motion of the crust and the liquid core.

Before the advent of Earth artificial satellites, these coefficients were determined by surface gravity measurements. Now the coefficients are determined by the perturbations they cause on satellite orbits as discussed below. One of the first discoveries of satellite geodesy, that the Earth is “pear shaped,” came from the orbital perturbation in eccentricity caused by $C_{30}$ [Section 5.4.4]. The recommended [2,227] model of the gravity field is (36,36), i.e. $n_{max}$=36 and $m_{max}$=36. Larger models are used for precision orbit calculations. High order models for Mars and Venus have obtained from numerous orbiting missions to these planets. Generally, accuracy of the coefficients decrease with increase in degree and order. Exceptions correspond to coefficients of potential terms that produced a resonance with the orbital motion of a particular satellite.

### 5.4.2 Non-spherical gravity perturbations

To apply the planetary equations to the non-central part of the field, write the gravity potential function as $V(r, \phi, \lambda) = -\frac{\mu}{r} - R(r, \phi, \lambda)$. Substitution of the complete **disturbing function** R into equations (5-18) generally has little practical value. The general approach is to divide the perturbations to the orbital elements into **secular** perturbations, **long period** perturbations, and **short period** perturbations. The secular variations result from averaging the equations of motion over one orbital period by assuming constant, mean values of the elements over that time. Recall the variation of parameter results of Section 5.2. The result generally is that some of the angular variables ($\Omega$, $\omega$ and $\lambda$) will change linearly with time. Inclusion of the slow change in these angular variables in the equations of motion produces the long period effects. When both the long period and secular effects are subtracted, only short period effects remain. These short period effects have periods no longer than the orbital period.

### 5.4.3 Oblateness Perturbations

As an example of this process, consider the $J_2$ term and the equation of motion for $\Omega$ for the elliptical orbit case

$$\frac{d\Omega}{dt} = -\frac{J_2\mu}{nab\sin i}\frac{\partial}{\partial i}\left[\frac{1}{r}\left(\frac{R}{r}\right)^2\left(\frac{3}{2}\sin^2\phi - \frac{1}{2}\right)\right]$$

since $P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$. To take the partial derivative r and $\phi$ must be replaced in favor of the orbital elements and time. Either the E or f forms could be used for r, so that r is only a function of time and the in-plane elements a, e and $\lambda$. For $\phi$, write $\sin\phi = \sin i \sin(\omega + f)$ from the law of sines, equation (1-1). Thus, the only dependence of this disturbing term on inclination is explicitly through sin(i). The final, exact equation of motion for $\Omega$ is

$$\frac{d\Omega}{dt} = -\frac{3\mu J_2}{nab}\frac{1}{r}\left(\frac{R}{r}\right)^2\cos i \sin^2(\omega + f) \tag{5-20}$$

None of the other five equations of motion are so easy to obtain. Since $J_2$>0, it is clear that $\dot{\Omega}$ is negative throughout the orbit if $i < 90^o$, vanishes at the equator as would be expected for a symmetric potential, and vanishes for polar orbits ($i=90^o$). The mean value of $\dot{\Omega}$ is the secular perturbation and is determined from the change in $\Omega$ over one orbital period assuming the orbital elements are constant on the right hand side. For this case, the independent variable in equation (5-20) is changed from time to f using $\frac{df}{dt} = \frac{h}{r^2}$ yielding

$$\frac{d\Omega}{df} = -\frac{3J_2R^2\cos i}{a^2(1-e^2)^2}(1+e\cos f)\sin^2(\omega+f)$$

The change in $\Omega$ in one orbit is obtained by integrating with respect to f from 0 to $2\pi$ to give

$$\Delta\Omega = \frac{-3\pi J_2\cos i}{(1-e^2)^2}\left(\frac{R}{a}\right)^2$$

Using the value of $J_2$ given above leads to about $0.5^o$ per orbit change for low inclination, LEO satellites. The secular rate per orbit is obtained by dividing by the period to give

$$\Delta\dot{\Omega} = \frac{-3nJ_2}{2(1-e^2)^2}\left(\frac{R}{a}\right)^2\cos i$$

yielding about $8^o$ per day change for low inclination, LEO satellites.

An alternate approach is to average the disturbing function over one orbital period before evaluating the partials on the right hand side of equations (5-18). The resulting disturbing function for the $J_2$ term is

$$\overline{R_2} = -\frac{1}{P}\int_0^P \frac{\mu J_2}{r}\left(\frac{R}{r}\right)^2\left(\frac{3}{2}\sin^2\phi - \frac{1}{2}\right)dt$$

Again switching to f as the independent variable yields after some algebra

$$\overline{R_2} = -\frac{\mu J_2}{2a(1-e^2)^{3/2}}\left(\frac{R}{a}\right)^2\left(\frac{3}{2}\sin^2 i - 1\right) \qquad (5\text{-}21)$$

Exercise 5-3. Perform the integration over one period to arrive at equation (5-21)

Thus the **mean disturbing function** for $J_2$ depends only on the orbital elements a, e and i. In view of equations (5-18) it is clear that there are no secular variations in a, e and i. Thus the mean orbit shape (a, e) is invariant and the mean inclination is constant. Physically this means that the average energy and z-component of angular momentum are preserved. The former should have been expected from the fact that the disturbing force is derivable from a potential function and the latter by the rotational symmetry of the potential due to $J_2$. The three secular variations due to $J_2$ are

$$\frac{d\Omega_s}{dt} = -\frac{3}{2}nJ_2\left(\frac{R}{p}\right)^2\cos i$$
$$\frac{d\omega_s}{dt} = \frac{3}{4}nJ_2\left(\frac{R}{p}\right)^2(5\cos^2 i - 1) \tag{5-22}$$
$$\frac{d\lambda_s}{dt} = \frac{3}{4}nJ_2\left(\frac{R}{p}\right)^2\sqrt{1-e^2}(3\sin^2 i - 2)$$

Exercise 5-4. Begin with equation (5-18) and equation (5-21) and verify the second of equations (5-22)

Interpretation of these equations shows that $\dot{\omega}_s = 0$ if $\cos i = \pm\frac{1}{\sqrt{5}}$, or i=63.43° or 116.56°. This angle is called the **critical inclination**. Since the argument of periapsis shows no secular variation at the critical inclination, the latitude of periapsis remains the same from orbit to orbit. The Molniya orbits in Section 7.3 are at the critical inclination so as to keep the periapsis at the latitude of the USSR. Below the critical inclination, periapsis regresses so that the time from one periapsis to the next is less than the “orbit period.” The last equation suggest that the mean motion is biased by $J_2$ since $\dot{M} = n + \dot{\lambda}$ similar to the phase change for the non-linear spring example in Section 5.2.

The perturbations discussed above are typical of those caused by all even zonal harmonics e.g. $J_4$, $J_6$, etc. When calculating accurate values for critical inclination or secular variations in $\Omega$, $\omega$ or $\lambda$, these additional terms must be considered.

Because of these types of perturbation to the orbit, a number of “periods” are in use. The **nodal period** is the time between successive ascending node passages. The **anomalistic period** is the time between successive periapsis passages based on the change in mean motion due to $J_2$ and other perturbations.

## 5.4.4 Odd-Zonal Perturbations

In a similar manner, the mean disturbing function for the third zonal harmonic is [4,349],

$$\bar{R}_3 = \frac{3\mu J_3 e}{8a}\left(\frac{R}{a}\right)^3(4\sin i - 5\sin^3 i)\left(1-e^2\right)^{-\frac{5}{2}}\sin\omega \tag{5-23}$$

Referring to equations (5-18) it is again seen that there is no change in mean energy due to $J_3$, but unlike $J_2$, there will be variations in inclination and eccentricity from the mean values. The eccentricity is particularly of interest because changes in eccentricity affect satellite lifetime. By direct substitution

$$\frac{de}{dt} = \frac{3}{8}nJ_3\frac{R^3}{ap^2}(5\sin^3 i - 4\sin i)\cos\omega \tag{5-24}$$

If all the terms on the right side were considered to be constant, this equation would suggest a secular variation in eccentricity. However, due to the $J_2$ effects, $\omega$ is varying linearly with time unless the orbital inclination is critical, i.e. $\cos^2 i=1/5$. Assuming that $\omega$ is the linear function of time given in equation (5-18) leads to the integral for the change in eccentricity from time $t_o$ to time t

$$\Delta e_3(t, to) = \frac{-J_3}{2J_2}\left(\frac{R}{a}\right)\sin i[\sin\omega(t) - \sin\omega(t_o)] \tag{5-25}$$

The maximum amplitude for the variation in eccentricity occurs for polar orbits. Since $J_3/J_2 \approx 0.002$ for the Earth, the maximum change in eccentricity for a LEO is about 0.001. This would produce a maximum variation in periapsis altitude of about 7 km with a period of $2\pi/\dot{\omega}$. For the Moon $J_2$=2.03x10$^{-4}$, so the node and periapsis for a low altitude lunar orbiter (LLO) will precess 1/5 of the rate per orbit as a LEO satellite. Also the lunar $J_3$=6x10$^{-6}$, yielding the ratio $J_3/J_2 \approx 0.03$ . Thus, the change in eccentricity due to $J_3$ is about 15 times larger. For LLO the effect of $J_3$ is a major consideration for orbit lifetimes.

Exercise 5-5. Derive equation (5-25) from equation (5-24).

There is also a long period variation in inclination which is of interest

$$\Delta i_3(t,t_o) = \frac{1}{2}\frac{J_3}{J_2}\left(\frac{R}{p}\right)e\cos i[\sin\omega - \sin(t_o)] \tag{5-26}$$

The perturbations discussed above are typical of those caused by all odd zonal harmonics e.g. $J_5$, $J_7$, etc. When calculating long term variations in i or e, these additional terms must be considered. Additional long period variations due to $J_3$ as well as secular, long period and short period terms for $J_2$ through $J_5$ are given in Koelle[8,8-26] through terms of order $J_2^2$.

### 5.4.5 Radiation pressure

Radiation pressure on an orbiting body occurs when photons strike the surface. These photons can be radiation directly from the sun (most of the energy is in the visible wavelengths), can be reflected from another body, or can be radiation, usually in the infrared, emitted from another body. Earth reflected radiation is particularly important for LEO satellites because of the high **albedo** of the Earth. The Earth has an albedo of about 0.3 at middle latitudes and 0.8 at the poles. Thus, between 30% and 80% of the incoming energy from the sun is reflected back into space. Modelling radiation pressure for a general perturbation approach is difficult because of shadowing, which produces a force that is a discontinuous function of orbit position. Radiation pressure is generally handled with special perturbations, i.e. numerical integration of the EOM. If shadowing is ignored, solar pressure can be analyzed using Lagrange's planetary equations as will be demonstrated below.

Solar pressure has been proposed as a propulsion system for a satellite. Solar sails can be constructed to "catch" photons and reflect them in a manner to produce thrust. This method needs a very high effective area to produce a substantial thrust.

Radiation from a body is normally specified in energy flux. For example, the energy flux from the Sun at 1 AU is about 1340 watts/m$^2$. The momentum flux, from which pressure can be calculated, is the energy flux divided by the speed of light. Hence, the solar momentum flux P is about 4.5x10$^{-6}$ N/m$^2$. The interactions of a photon with a surface ranges from passing through without any absorption (transparent material), having a probability greater than zero of being absorbed (translucent), completely absorbed (black body) and reflected (mirror). The solar pressure $p_s$ is modeled as

$$p_s = \alpha P \tag{5-27}$$

where $0 \le \alpha \le 2$. Transparent materials have $\alpha$=0 and mirrors have $\alpha$=2, i.e. transparent materials absorb none of the momentum flux and mirrors reverse the direction of the momentum flux, effectively reacting to twice the incoming flux. The total force is obtained by integrating equation (5-27) over the exposed area of the body. Note that in addition to producing a net force on the satellite, solar pressure can also produce significant torques on unsymmetrical satellites which must be considered for attitude control.

Exercise 5-6. Show that a 100 kg satellite with a cross sectional area of 2 m$^2$ and albedo of 1 would experience about 1 micro-g of acceleration due to radiation pressure.

Now consider a spherical satellite of the Earth that is not passing through the shadow of the Earth. Assume a homogeneous reflecting surface so that the solar pressure is constant and away from the Sun. The radiation force is therefore $-\alpha P A \mathbf{e}_s$, where $\mathbf{e}_s$ is the unit vector from the central body to the Sun and A is the effective cross sectional area. Assuming that the Sun does not move over one orbital period, this is a constant force and hence derivable from the **disturbing function** $R = -\alpha P A \mathbf{e_s} \cdot \mathbf{r} = \beta \mathbf{e_s} \cdot \mathbf{r}$ where $\mathbf{r}$ is the position vector and in terms of the orbital elements is given by equation (3-21). The average of R over an orbit can be obtained using equations (3-26) and (3-27) to yield the disturbing potential for long period and secular variations $\bar{R} = \kappa a \mathbf{e_s} \cdot \mathbf{e}$ where $\mathbf{e}$ is the eccentricity vector and $\kappa = -3\beta/2$. Hence, the eccentricity and the angle between the direction to the Sun and the major axis of the ellipse completely determine the perturbations.

Referring to the planetary equations (5-18), it is seen that there is no secular variation in energy because the work done by solar pressure as the satellite approaches the sun is equal to the work done as the satellite recedes from the sun. This conclusion is not generally valid if shadowing occurs during the orbit. From a mission design viewpoint the most interesting variation is in the eccentricity, where solar pressure can produce a long term variation in eccentricity much like the $J_3$ effect. Only a special case will be demonstrated here. Let the Sun be on the x-axis, then from equation (3-33) $\bar{R} = a\kappa(\cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i)$. The two elements of interest are $\omega$ and e.

Evaluating equations (5-18) for these two variables and then letting inclination become zero so that the Sun is in the orbit plane yields

$$\dot{e} = \frac{\kappa b}{na^2}\sin(\omega + \Omega) \qquad \dot{\omega} = \frac{\kappa b}{nea^2}\cos(\omega + \Omega)$$

These equations show that if solar pressure is the only perturbing force, the argument of periapsis will precess until $\cos(\omega+\Omega)=0$ and the precession is very rapid for small eccentricity. When the sun is along the semi-latus rectum, the precession stops, then e will either increase until the satellite hits the planet or decrease until a circular orbit is achieved. Though the equations here are not applicable when e=0, it can be shown that a circular orbit will become increasingly elliptical with the major axis at right angles to the sun line. For most satellites the $J_2$ secular variation in $\Omega$ and $\omega$ will dominate the solar pressure precession in $\omega$, so the eccentricity will undergo a long period variation. Recall again that these results are for the no shadowing case.

Exercise 5-7. Use the toolbox to plot the 24 hour ground track of the LEO satellite a=7000, e=0.05, $i=55^o$, $\Omega=60^o$, $\omega=45^o$, $\tau$=July 4, 2000, 13 hrs, 55 min, 34.56 sec. Compare tracks with and without $J_2$ precession.

## 5.5 Gauss' Form of the Perturbation Equations

Drag and some other forces can not be formulated as potential functions. It is therefore of interest to have the analog of the planetary equations in a form where the perturbing force or acceleration appear explicitly. There are numerous approaches to arrive at these equations. The direct method is to start with equation (5-10) and substitute the results from equations (5-13) and (5-14) directly into $\left[\frac{\partial r}{\partial \mathbf{c}}\right]$. Another method is to consider the effect of an impulse or instantaneous change in $\mathbf{v}$ applied at some point in the orbit. Since the motion before and after the impulse is pure two body motion, the change in the elements across the impulse can be obtained by applying differential calculus to any two body equation. For example, from the vis-viva integral

$$2\mathbf{v} \cdot \delta\mathbf{v} = \mu\frac{\delta a}{a^2} \qquad \Rightarrow \qquad \delta a = \frac{2a^2}{\mu}\mathbf{v} \cdot \delta\mathbf{v}$$

where the fact that an impulse does not change $\mathbf{r}$ has been utilized. Interpretation of this equation shows that the most effective location in orbit to change the energy is to apply an impulse at periapsis (where v has the maximum value) along the velocity vector. Energy is not affected by an impulse normal to the velocity vector. If instead of an impulse, it is assumed that the $\delta\mathbf{v}$ occurred over a finite but small time $\delta t$, then taking the limit leads to

$$\frac{da}{dt} = \frac{2va^2}{\mu}a_t$$

where $a_t$ is the tangential component of the disturbing acceleration. This is an exact equation of motion.

The complete set, given below, is from [5,489] for the case where the perturbation acceleration is projected along orthogonal axes that are along the orbit tangent ($a_t$), normal to the orbit plane ($a_h$) and normal to the velocity vector in the orbit plane ($a_n$). In the Euler form of the perturbation equations, it is not convenient to utilize $\lambda$ or $\tau$ as an orbital element for reasons discussed in [5]. Instead the last equation is given for M explicitly. Danby [4] and many other books present the equations of motions for perturbations that are projected in the radial, normal to the orbit, and circumferential directions.

$$
\begin{aligned}
\frac{da}{dt} &= \frac{2va^2}{\mu}a_t \\
\frac{de}{dt} &= \frac{1}{v}\left(2(e+\cos f)a_t - \frac{r}{a}\sin f a_n\right) \\
\frac{di}{dt} &= \frac{r\cos\theta}{h}a_h \\
\frac{d\Omega}{dt} &= \frac{r\sin\theta}{h\sin i}a_h \\
\frac{d\omega}{dt} &= \frac{1}{ev}\left[2\sin f a_t + \left(2e + \frac{r}{a}\cos f\right)a_n\right] - \cos i\frac{d\Omega}{dt} \\
\frac{dM}{dt} &= n - \frac{b}{eav}\left[2\left(1+\frac{re^2}{p}\right)\sin f a_t - \frac{r}{a}\cos f a_n\right]
\end{aligned}
\tag{5-28}
$$

Interpretation of equations (5-28) can lead to an understanding of where to apply impulses to achieve maximum change in an orbit parameter. Changing energy or semi-major has already been discussed. Changing orbital inclination is often a mission requirement. It is seen that only an impulse normal to the orbit plane will change inclination and that the most effective location is where $r\cos(\omega+f)$ is a maximum/minimum. The cosine reaches an extremum when the satellite is on the node line, so this is the most efficient location for a circular orbit. For the high eccentricity transfer orbits from LEO to GEO the optimal location for a single impulse would clearly be near apoapsis. Similar arguments can be made for $\Omega$, but the other variables are not so obvious. Since there are three components to the impulse, at most three elements can be controlled with one maneuver. Of course, the remaining three elements may also change. For a particular orbit, the optimal location for performing an impulse can be formulated as a constrained optimization problem and the solution found by searching numerically around the orbit.

### 5.5.1 Drag

Any planetary atmosphere experienced by an orbiting body will cause drag and perhaps other forces and moments on the satellite. In the free molecular flow region that is usually associated with satellites with more than a few orbits of lifetime, side forces are generally negligible. Drag transfers kinetic energy of the satellite to thermal energy of the atmosphere. Since satellite energy is decreasing, the semi-major axis for elliptic motion must be decreasing and mean motion is increasing. This leads to the seemingly paradoxical statement that “drag makes the satellite move faster.” The drag acceleration **d** is generally modeled as

$$\mathbf{d} = -\frac{1}{2}\rho v\left(\frac{C_d A}{m}\right)\mathbf{v} \tag{5-29}$$

where A is the satellite reference area, m is the mass, $C_d$ is the drag coefficient, **v** is the velocity vector, v is the speed, ρ is the atmospheric density, and $C_d A/m = \beta$ is the **ballistic coefficient.** Atmospheric density can vary with altitude, planet-sun distance, day/night, latitude, local solar time, solar activity, etc. and models of these variations are not precise. As a result the analyst must be careful in modeling drag phenomena. Letting A be the cross-section of the spacecraft exposed to free molecular flow, $C_d \geq 2$. $C_d$=2 if the linear momentum of all incoming molecules is completely absorbed by the satellite. This situation occurs if the satellite surface has a **momentum accommodation coefficient** of unity. However, these gas-surface interactions are very complicated. Simple models assume that some fraction of the incoming molecules are not absorbed by the surface and that most absorbed molecules are quickly emitted from the surface after coming into thermal equilibrium with the surface. This generally leads to $C_d \approx 2.2$.

Drag is generally the dominate force that defines a satellite's orbital lifetime and requires propulsive capability for orbit maintenance. The definitive study of drag effects is given by King-Hele [6]. On the other hand, atmospheric drag has been used for **aerobraking** which is the process of reducing orbital energy to a desired level by dipping into an atmosphere. This process can significantly reduce propulsive requirements. In any case, if latitudinal and longitudinal variations in density are significant, the usual approach is to numerically integrate the equations of motion in rectangular coordinates using equation (5-29) for the perturbing force. If such variations are negligible or to gain insight into the effects of drag on orbit parameters, density can be assumed to only be a function of altitude.

From equations (5-28) it is seen that only a, e, ω and M are perturbed by a tangential drag force. For orbit lifetimes, the perturbations to a and e are of particular interest because $r_p$=a(1-e). Referring to the first two of equations (5-27), change the independent variable from time to eccentric anomaly using Kepler's equation (3-14) and use the vis-viva integral (3-5) to write $v^2 = \frac{\mu}{a}\left(\frac{1 + e\cos E}{1 - e\cos E}\right)$ to finally yield

$$\frac{da}{dE} = -\beta a^2 \rho \frac{(1 + e\cos E)^{3/2}}{(1 - e\cos E)^{1/2}} \qquad \frac{de}{dE} = -\beta p \rho \cos E \sqrt{\frac{1 + e\cos E}{1 - e\cos E}} \tag{5-30}$$

If density is modeled as only a function of altitude, density can be written as a function of E. Further, if a and e can be assumed to be constant during a single pass through the atmosphere, these equations can be integrated numerically to yield the change in a and e during one orbit.

Exercise 5-8. Derive equations (5-30) following the directions above.

Numerous approximations [6] have been made to obtain analytic solutions to these equations. First note that, omitting density, the remaining terms are periodic, even functions of E and can be expanded in a Fourier cosine series in E, for example

$$\cos E\sqrt{\frac{1+e\cos E}{1-e\cos E}} = \left(\frac{e}{2}+\ldots\right)+\left(1+\frac{3}{8}e^2+\ldots\right)\cos E+\left(\frac{e}{2}+\ldots\right)\cos 2E+\ldots$$

The coefficients are infinite power series in eccentricity. For analytic solutions, atmospheric density is modeled by an exponential in altitude, i.e.

$$\rho = \rho_o \exp\left(-\frac{h-h_o}{H_s}\right) \tag{5-31}$$

where $\rho_o$ is the density at reference altitude $h_o$ and $H_s$ is the **density scale height**. For the Earth, $H_s$ may range from 30 to 100 km depending on altitude, solar cycle, and other geophysical parameters. The reference altitude is usually taken as the periapsis altitude, $h_o=h_p$. To utilize this model for density in equations (5-30), note that $h-h_p = r-r_p = ae(1-e\cos E)$. With this substitution and the Fourier series expansions, the right hand sides can be integrated over one orbit. The results are infinite series of Bessel functions with coefficients that are infinite series in eccentricity[Reference 6, Chapter 4]. Reference 6 provides numerous approximations to equations (5-30) based on the values of e and $\alpha=ae/H_s$. Only one of the expansions is given here as it applies to many LEO satellites and is applicable if $0.02<e<0.2$ and $\alpha>3$

$$\Delta a = -2\pi\beta a^2\rho_p\exp(-\alpha)\left[I_o+2eI_1+\frac{3}{4}e^2(I_o+I_2)+\frac{1}{4}e^3(3I_1+I_3)+O(e^4)\right] \tag{5-32}$$

where the argument of the modified Bessel functions is $\alpha$, i.e. $I_n=I_n(\alpha)$.

### 5.5.2 N-Body Perturbations

To obtain the equation of relative motion of $m_1$ with respect to $m_2$ including the effects of the remaining n-2 bodies, subtract equation (2-10) with i=2 from the same equation with i=1,

$$\ddot{\mathbf{r}} = -\frac{G(m_1+m_2)\mathbf{r}}{r^3} - G\sum_{j=3}^{n}\left[\frac{\mathbf{r}_{1j}}{r_{1j}^3}-\frac{\mathbf{r}_{2j}}{r_{2j}^3}\right] \tag{5-33}$$

Even though the perturbation term on the right is derivable from a disturbing function, obtaining an average disturbing function, to study secular and long period variations, is difficult because the other bodies are in motion. Prior to the development of high speed computers, these equations were used as the basis for planetary theories. Generally, the approach is to consider a single disturbing body at a time. Double averaging is done over the two orbital periods to obtain the averaged disturbing function. The interested reader can consult Reference 9 for details. For LEO satellites, secular and long period variations due to the Moon and Sun are several orders of magnitude smaller than the $J_2$ secular terms, but can be substantial for high eccentricity orbits.

## 5.6 Special Perturbations

Solving the equations of motion using numerical integration techniques is called the method of **special perturbations**. When choosing an integration scheme to numerically integrate equation (5-1), speed, accuracy, storage, and complexity must all be addressed. The main consideration is the selection of a method for numerical integration. The most efficient methods are second order, multi-step, constant step size which are very efficient for low to moderate eccentricity orbits. Reference 4 provides an introduction to the numerical procedures used in orbital mechanics including interpolation, extrapolation, differentiation and integration.

## 5.7 Problems

5-1. Apply the planetary equations to equation (5-23) to derive equation (5-25).

5-2. Apply the planetary equations to equation (5-23) to derive equation (5-26)

5-3. Derive an expression for the first order $d\omega/dt$ due to $J_3$ similar to equation (5-22).

## 5.8 Astronautics Toolbox

1. Develop a function [wdot, Wdot, Lamdot]=J2Precess(a,e,i,J2,mu,R) that will return the precession rates in radians per unit of time.
2. Modify the Orb2X routine (Section 3.14) so that $\dot{\Omega}$, $\dot{\omega}$ and $\dot{\lambda}$ due to $J_2$ are included to change $\Omega$, $\omega$ and $\lambda$ in the orbit propagation. Call the new routine Orb2XJ2 and add J2 and R to the input set. Use J2Precess.
3. Write a function that will plot ground tracks for a LEO satellite given the orbital elements in the J2000 equatorial system, (Section 1.2.2). Time of periapsis is in ymdhms format (Section 1.6) and time interval for the plot starts at periapsis and stops an input time in days later. Include an option for turning $J_2$ precession on or off. An option is to also limit applicability to years near 2000 so that precession of the vernal equinox does not have to be considered in the terrestrial longitude calculation using sidereal time (Section 1.3.5). Make maximum use of existing toolbox functions. It is recommended that a considerable design effort precede implementation of this function.
4. Develop a function that will integrate equations (5-30) over one orbital period. Assume an exponential atmosphere. [da,de]=dragI(a,e,Cd,A,m,rho0,h0,Hs,ichk).
5. Develop a function, [da,de]=dragKH(a,e,Cd,A,m,rho0,h0,Hs,ichk), that will evaluate equation (5-32)
6. Write a test program that will compare the relative error in da between dragKH and dragI over the applicable range of eccentricities using a=6800 km and Hs=40 km.

## 5.9 References


1. Abramowitz, M. and Stegun, I., **Handbook of Mathematical Functions**, Dover Publications Inc., 1965.
2. Seidelmann, K., **Explanatory Supplement to the Astronomical Almanac**, University Science Books, Mill Valley, CA, 1992.
3. Fitzpatrick P.M., **Principles of Celestial Mechanics**, Academic Press, Inc., 1970

4. Danby, J.M.A., **Fundamentals of Celestial Mechanics**, Willmann-Bell, Richmond, Va., 1989.
5. Battin, R.H., **An Introduction to the Mathematics and Methods of Astronautics**, AIAA Education Series, 1987.
6. King-Hele, D., **Theory of Satellite Orbits in an Atmosphere**, Butterworths, 1961.
7. Moulton, F.R. **An Introduction to Celestial Mechanics**, The MacMillan Company, New York, 1914. Also available from Dover Press.
8. Koelle, H.H., **Handbook of Astronautical Engineering**, McGraw-Hill, 1961.
9. Brouwer, D. and Clemence, G., **Methods of Celestial Mechanics**, Academic Press, 1961.

# Chapter 6 - Orbit Transfer and Powered Flight

## 6.1 Introduction

The general term **orbit transfer** refers to the maneuvers required to change the orbit of a spacecraft from an initial orbit to a terminal orbit. A propulsive maneuver is called **impulsive** if the maneuver time is very short compared to some orbit characteristic time, e.g. period. For impulsive maneuvers the change in position during the maneuver is neglected. Otherwise, the maneuver is termed **continuous** or **finite burn**, for example low thrust orbit transfers. In this chapter the emphasis will be on impulsive orbit transfers. Often "**orbit transfer**" is used in the more limited sense where only the five spatial elements (a, e, i, $\Omega$ and $\omega$) are specified for the initial and final orbits. In this case the times of departure and arrival are unimportant**.** **Interception** requires that the transfer orbit pass through a specific moving point in space. For interception, individual values of the orbital elements are not important, but time optimal and fuel optimal transfers are of interest. **Rendezvous** specifies all six of the orbital elements of the terminal orbit, thereby assuring coincidence in both position and velocity. Optimizing orbit transfers involves minimizing propulsive requirements, transfer time, or other mission parameters often subject to equality or inequality constraints. This chapter provides the formulation for various orbit transfers. Powered flight in this chapter will be limited to elementary launch considerations including structural mass, propellant mass and staging.

## 6.2 Powered Flight

Powered flight is normally thought of as being based on the expulsion of some mass from the vehicle to produce a change in velocity of the vehicle. Chemical rocket motors, cold gas jets, ion rockets, and rail guns are examples. The resulting change in vehicle velocity is a result of Newton's Second Law [Section 2.2.1], which states that the time rate of change of linear momentum is equal to the sum of the external forces.

### 6.2.1 Rocket equation

Let the linear momentum of the entire vehicle at time t be m**v**, where m is the mass at time t and **v** is the velocity of the center of mass of the rocket with respect to an inertial coordinate frame. The change in m**v** is due to the decrease in mass of the rocket when propellant, and effectively mass, is expelled from the vehicle. Define the mass flow rate as $\dot{m} < 0$, which is taken as negative since the mass of the rocket is decreasing. To apply Newton's Second Law to this system, the total linear momentum is evaluated at time t and t+$\Delta$t. At time t, the linear momentum is m**v**, while at time t+$\Delta$t, the linear momentum is the sum of the momentum of the rocket (with decreased mass) and the momentum of the expelled mass or

$$(m + \dot{m}\Delta t)(\mathbf{v} + \Delta\mathbf{v}) + (-\dot{m}\Delta t)(\mathbf{v} + \mathbf{c})$$

where **c** is the **exhaust velocity** with respect to the rocket.

The change in linear momentum is equal to the total impulse due to external forces:

$$(m + \dot{m}\Delta t)(\mathbf{v} + \Delta\mathbf{v}) - (\dot{m}\Delta t)(\mathbf{v} + \mathbf{c}) - m\mathbf{v} = \mathbf{F}_{ext}\Delta t$$

Simplifying yields:

$$m\Delta v + (\dot{m}\Delta t)(\Delta\mathbf{v} - \mathbf{c}) = \mathbf{F}_{ext}\Delta t$$

Dividing by $\Delta t$ and taking the limit as $\Delta t \to 0$:

$$m\frac{d\mathbf{v}}{dt} = \mathbf{c}\frac{dm}{dt} + \mathbf{F}_{ext} \tag{6-1}$$

Equation (6-1) is known as the **rocket equation**, which describes the acceleration of the rocket due to **thrust** (=$\mathbf{c}$ dm/dt) and external forces.

A comparative measure of the performance of a rocket is the **specific impulse**, which is defined as the ratio of the thrust divided by the propellant flow rate in weight/second at sea level:

$$I_{sp} = \frac{\dot{m}c}{\dot{w}} = \frac{\dot{m}c}{\dot{m}g_o} = \frac{c}{g_o} \tag{6-2}$$

A table of specific impulses for various commercial engines is shown below in Table 6-1 (See Reference 2).

Table 6-1. Typical Rocket Motor Characteristics

| Engine | Thrust (lbf) | Fuel | Oxidizer | Isp |
|---|---|---|---|---|
| Rocketdyne RS-27 (Delta) | 207000 | RP-1 | LO2 | 262 (S.L.) |
| Aerojet AJ110 | 9800 | UDMH/N2H4 | N2O4 | 320 (Vac) |
| TRW TR-201 (Delta) | 9900 | UDMH/N2H4 | N2O4 | 303 (Vac) |
| United Technologies Orbus 6 | 23800 | Solid | | 290 (Vac) |
| United Technologies Orbus 21 | 58560 | Solid | | 296 (Vac) |
| Pratt & Whitney RL-10 | 16500 | LH2 | LO2 | 444 (Vac) |

In the rest of the chapter, the scalar version of equation (6-1) will be utilized by assuming that **c** and **v** are parallel and unchanging in direction. A simple solution to equation (6-1) occurs when there are no external forces, such as atmospheric drag or gravity, acting on the rocket ($F_{ext}$=0):

$$m\,dv = c\,dm \quad \text{or} \quad dv = c\frac{dm}{m}$$

This equation can be integrated assuming that the exhaust velocity c is constant yielding the **ideal velocity equation**:

$$\Delta v = v_t - v_o = c\log\frac{m_o}{m_t} = I_{sp}g_o\log\frac{m_o}{m_t} \tag{6-3}$$

where $\Delta v$ is called the **characteristic velocity** of the rocket when t is the final or burnout time.

Exercise 6-1. For each rocket motor in Table 6-1, calculate the ratio of final to initial mass for a velocity increase of 5000 m/s. Plot and interpret results.

In the presence of external forces, equation (6-1) can be rewritten as:

$$dv = c\frac{dm}{m} + \frac{F_{ext}}{m}dt$$

Under certain forms of the force, the solution can be written as

$$v_t - v_o = c \ln\frac{m_o}{m_t} + \int_{t_o}^{t} \frac{F_{ext}}{m}dt \tag{6-4}$$

For rockets in vertical ascent through a constant Earth gravity field, equation (6-4) reduces to:

$$v_t - v_o = g_o I_{sp} \ln\frac{m_o}{m_t} - g_o(t - t_o) \tag{6-5}$$

So it is seen that the **gravity loss** term, $g_o(t-t_o)$ can only be neglected if the burn time is short compared to the specific impulse. Orbit trim maneuvers and launch vehicles with high thrust to weight ratios (meteorological rockets, fireworks) often fall into this category. In vehicles with low thrust to weight ratios (most scientific, manned and commercial launches), the gravity term cannot be ignored.

The $I_{sp}$ is primarily determined by the fuel selection (Table 6-1). So the velocity gained can only be increased by increasing the ratio of initial mass to final mass. The initial mass is the sum of the fuel mass, the structural mass, and the payload mass

$$m_o = m_f + m_s + m_p \tag{6-6}$$

and the final mass is $m_t=m_s+m_p$. A significant parameter is the **structural mass factor**

$$\beta = \frac{m_s}{m_s + m_f} \tag{6-7}$$

which is a measure of the structural efficiency of the vehicle. Solid rockets generally have lower $\beta$ values than liquid systems, but at the cost of lower $I_{sp}$. A value of $\beta$ as low as 0.04 (Atlas launch vehicle) would be considered a very efficient structural design. Values in the range of 0.08 to 0.12 are more typical. So even with no payload and ignoring gravity losses and drag, the maximum velocity for a single stage vehicle with $\beta$=0.1 and $I_{sp}$= 300 is less than 7000 m/s. This is less than the orbital speed for a LEO of 7500 m/s. Consequently staging of launch vehicles is required to reach LEO speeds when external forces and non-vertical launches are considered.

Exercise 6-2. Neglecting gravity losses and with payload mass of zero, calculate the structural mass fraction for each engine in Table 6-1 required to provide LEO orbital speed of 7500 m/s.

Exercise 6-3. For a payload of 500 kg, structural mass fraction of 0.08, and final velocity of 7000 m/s, calculate the lift off mass for a rocket powered by the Pratt & Whitney engine. What is the burn time (careful of units)? Would it be reasonable to ignore gravity losses for a vertical flight?

### 6.2.2 Staging

Consider the case of n-stages. The payload at the end of stage i is approximated by the initial mass for stage i+1. The final mass of the vehicle is the payload mass plus the structural mass of the last stage. The overall payload to initial mass has been increased by jettisoning the structural weight of the first n-1 stages. Ignoring drag and gravity losses, the velocity change for a **multistage** rocket with N stages is the sum of the velocities of the individual stages

$$\Delta v = \sum_{i=1}^{N} c_i \ln\left\{\frac{m_{o,i}}{m_{o,i} - m_{f,i}}\right\} \tag{6-8}$$

For the case where $\Delta v$ and $\beta_i$ are fixed, there is a choice of how to distribute the total mass among the stages. An optimization problem associated with rocket staging is to minimize the overall mass to payload ratio

$$\frac{m_o}{m_p} = \prod_{i=1}^{N}\left\{\frac{m_{o,i}}{m_{o,i} - m_{f,i} - m_{s,i}}\right\} \tag{6-9}$$

subject to the constraint that the sum the characteristic velocities be a specified value, $\Delta v_T$:

$$\Delta v_1 + \Delta v_2 + \ldots + \Delta v_N - \Delta v_T = 0 \tag{6-10}$$

This problem, which can be solved using the Lagrange multiplier process [1], is of little practical value because of the assumptions of straight line flight and the omission of drag and gravity effects. Rather, optimal multistage launch trajectory analyses are performed using sophisticated models that include the variation of g with altitude, inclusion of Earth rotation, aerodynamics, pitch and yaw of the thrust vector, and many other factors. Nevertheless, the optimal solution can be utilized to estimate the upper bound for performance.

## 6.3 Impulsive Maneuvers

Midcourse corrections and orbit transfers maneuvers are often performed with thrust durations that are very short compared to the orbital period and with the thrust applied in one direction. Such maneuvers, at least for purposes of mission design, can be approximated by impulses. During an impulse, the position of the vehicle is assumed to not change, but the velocity undergoes an instantaneous change in magnitude and/or direction. Most orbit changes must consider the three dimensional nature of the problem and conditions for optimality are often difficult to develop. Three and four impulse optimal transfers are still a research topic. The two dimensional case, on the other hand, admits to some analytical solutions. Finally, optimal transfers are often compromised by launch windows, limitations on the available sizes of solid rockets boosters, and other missions requirements [4,137].

**In-plane** or **coplanar** orbit changes are those maneuvers that do not change the direction of angular momentum, but may change a, e, $\omega$, and/or $\tau$. The $\Delta v$ for such a maneuver must therefore be in the orbit plane. Several different transfers will be discussed here, involving various conic sections and impulses. The most general planar case, optimal transfer between two arbitrary coplanar elliptical orbits, requires finding the roots of an eighth degree polynomial [3,525] just to evaluate the necessary conditions for optimally. Constraints can be placed on the possible transfer orbits [3,237] but often zeroth order analyses are used to find the optimum.

### 6.3.1 Single impulse transfers.

A single impulse transfer between two orbits can only be considered if the two orbits have at lease one common point in space. The transfer must then occur at one of the intersection points. For small impulsive changes in the orbital elements, Gauss' form of the perturbation equations (5-27) can be utilized to solve for all three components of the impulse. In general, only three orbital elements can be controlled since there are only 3 components to an impulse. Also some combinations of changes may be impossible. For example, from the third and fourth of equations (5-27) it is clear that inclination can not be changed without changing the nodal location unless the intersection point is on the line of nodes. In this case the maneuver must be performed on the line of nodes, i.e. $\sin\theta=0$. Typical single impulse maneuvers include

1. maneuvers at apoapsis to raise periapsis to reduce drag effects, requires $\delta a$ and $\delta e$ changes,
2. maneuvers to rotate orbit in-plane to place periapsis over a specified latitude, $\delta\omega$ change,
3. maneuvers at periapsis to change orbital period, $\delta a$ and $\delta e$ change,
4. maneuvers near the line of nodes to change inclination.

Two impulse maneuvers, to accomplish the same orbit changes, often require much less $\Delta v$ than the single impulse. For example, the optimal two impulse maneuver to change $\omega$ requires half of the single impulse $\delta v$ [4, 109].

### 6.3.2 Two-impulse transfer between coplanar circular orbits

Consider the transfer of a spacecraft in a circular orbit of radius $r_1$ to a coplanar circular orbit of radius $r_2$, as shown in Figure 6.1. Assuming no external forces during the transfer, the transfer orbit is a conic. The first impulse, $\Delta\mathbf{v_1}$, places the spacecraft on the transfer orbit and the second, $\Delta\mathbf{v_2}$, recircularizes at $r_2$. The periapsis of the transfer orbit cannot lie outside the inner orbit, and the apoapsis must lie outside or be tangent to the outer orbit. The transfer orbit must therefore satisfy the conditions:

$$\begin{aligned} r_p &= a(1-e) \le r_1 \\ r_a &= a(1+e) \ge r_2 \end{aligned} \tag{6-11}$$

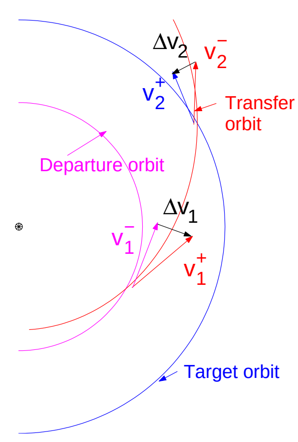



Figure 6-1. Two impulse transfer between coplanar circular orbits

The semi-major axis and the eccentricity provide constraints on the possible range of $v_1$ and $v_2$. Here $r_p$ and $r_a$ are selected as independent variables in lieu of a and e.

The four relevant velocities can be obtained from the vis-viva equation (Equation (3-5)). Before the first burn the spacecraft has circular speed from the inner orbit:

$$v_1^- = \sqrt{\frac{\mu}{r_1}} \,. \tag{6-12}$$

The transfer orbit speed immediately after the first burn is

$$v_1^+ = \sqrt{2\mu\left(\frac{1}{r_1} - \frac{1}{r_a + r_p}\right)} \tag{6-13}$$

Before the second impulse

$$v_2^- = \sqrt{2\mu\left(\frac{1}{r_2} - \frac{1}{r_a + r_p}\right)} \tag{6-14}$$

and afterwards

$$v_2^+ = \sqrt{\frac{\mu}{r_2}} \tag{6-15}$$

The angle between the two vectors $v_1^+$ and $v_1^-$ is the **flight path angle** on the transfer conic $\gamma_1$, and can be obtained from the angular momentum $h=r_1 v_1^+ \cos\gamma_1 = \sqrt{\frac{2\mu r_p r_a}{r_p + r_a}}$. From the law of cosines, the speed change required to transition to the transfer orbit is

$$\Delta v_1 = \sqrt{v_1^{+2} + v_1^{-2} - 2v_1^+ v_1^- \cos\gamma_1} \tag{6-16}$$

which is a function of $r_1$, $r_p$, and $r_a$. Likewise, for the second impulse

$$\Delta v_2 = \sqrt{v_2^{+2} + v_2^{-2} - 2v_2^+ v_2^- \cos\gamma_2} \tag{6-17}$$

is a function only of $r_2$, $r_p$, and $r_a$. Equation (6-16) and (6-17) are equally applicable if $r_2<r_1$ and if the maneuver occurs at either of the two intersection points of the transfer orbit with the initial orbits. Elliptic transfer orbits that make less than one complete orbit about the primary are classified as four **Types**. **Type I:** during the transfer the true anomaly satisfies $0 \le f \le \pi$. That is, the motion is away from the primary during the entire transfer. **Type II:** periapsis is not contained in the transfer arc but apoapsis is. **Type III:** $\pi \le f \le 2\pi$ and **Type IV:** the transfer includes periapsis but not apoapsis. These definitions are not uniformly accepted in the community.

Figure 6-2 provides the $\Delta v$ requirements for transfer from a LEO at a=6700 km to a geosynchronous orbit as a function of $r_a$ and $r_p$ of the transfer ellipse. The blue lines are the $\Delta v$ for the first impulse required to establish the transfer orbit. These velocities are sensitive to $r_p$ because reducing $r_p$ requires turning as well as increasing the velocity with the first impulse. They are a weaker function of $r_a$ because near periapsis a small change in velocity provides a large increase in semi-major axis.The magenta lines are for the second impulse required to circularize the transfer orbit at geosynchronous altitude. These lines show the opposite sensitivity. Increasing $r_a$ provides significant sensitivity because the velocity on the transfer orbit at $r_2$ increase in magnitude and flight path angle.

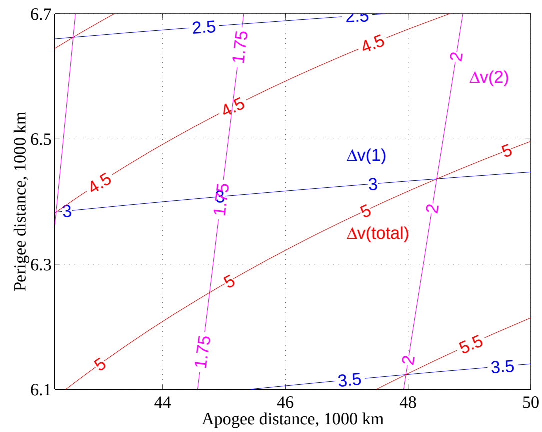



Figure 6-2. Delta V requirements for transfer from $r_p$=6700 km to geosynchronous orbit.

Figure 6-3 shows the total $\Delta v$ and the transfer time in hours for the transfers in Figure 6-2. Of course, transfer time is completely determined by the first maneuver. For a fixed total $\Delta v$, the quickest transfer occurs when $r_p=r_1$ or $\gamma_1=0$. With $r_p=r_1$, it takes almost 500 m/s to reduce the flight time from 100 hrs. to 70 hours.

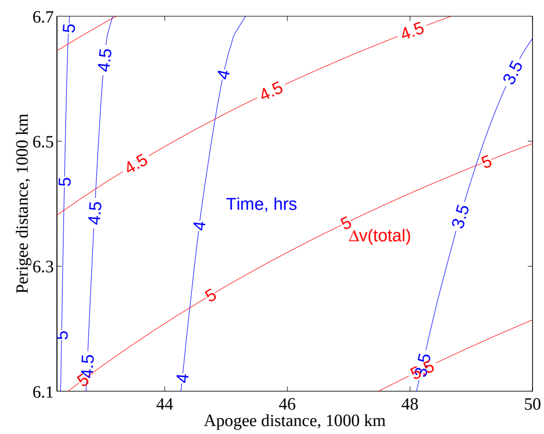



Figure 6-3. Transfer time and total delta V for transfer from $r_p$=6700 km to GEO.

## 6.3.3 Hohmann transfer

From equations (6-16) and (6-17) it can be shown [1] that minimum $\Delta v = \Delta v_1 + \Delta v_2$ occurs when $\gamma_1$ and $\gamma_2$ are zero. This minimum speed change corresponds to a **minimum fuel** transfer. The transfer ellipse is tangent to both circular orbits. This ideal maneuver is known as a **Hohmann transfer** after Walter Hohmann who in 1925 hypothesized that such a maneuver would be optimal. The semi-major axis of the transfer orbit is $a_H = \frac{r_1 + r_2}{2}$, the transfer time $T_H = \pi\sqrt{a_H^3/\mu}$ is half of the period of the ellipse, and the eccentricity can be written in terms of the radii, $e_H = (r_2 - r_1)/(r_1 + r_2)$. Equations (6-16) and (6-17) reduce to

$$\Delta v_1 = \sqrt{\frac{\mu}{r_1}}\left\|\sqrt{\frac{2r_2}{r_1 + r_2}} - 1\right\| \qquad \Delta v_2 = \sqrt{\frac{\mu}{r_2}}\left\|\sqrt{\frac{2r_1}{r_1 + r_2}} - 1\right\| \tag{6-18}$$

For the example in Figure 6-2 and Figure 6-3, the Hohmann transfer occurs in the upper left corner. The corresponding values are $\Delta v_1$=2.42 km/s, $\Delta v_2$=1.46 km/s, and the transfer time is 5.3 hrs.

Exercise 6-4. Starting with equation (6-16) and equation (6-17) verify Equations (6-18).

While the Hohmann transfer is the most economical transfer in terms of fuel usage, there are a number of reasons that they are not practical. First, the transfer time is significantly longer than some slightly less energy efficient transfer orbits. Hohmann transfers may also be physically impossible, for example, Hohmann transfers from an intermediate LEO to geosynchronous equatorial orbit are only possible if the LEO is in the equatorial plane. This will normally require a launch site near the equator. Hence, establishing GEO orbits using Hohmann transfers from non-equatorial launch sites requires a “dog leg” launch trajectory or an orbit transfer with a significant orbit plane change to establish the equatorial orbit. Similarly, because the planetary orbits are not in the same plane, Hohmann transfers are generally not possible for interplanetary transfers. Near Hohmann transfers, i.e. near 180 deg transfers with small plane changes, are possible if departure and arrival occurs near the line of nodes of one orbit with respect to the other.

If the planets were in circular orbits and in the same plane, the 180 degree transfer places strict conditions on the relative angular position of the departure and arrival planets at the time of departure. Since the transfer time and transfer angle are fixed, the angular position of the target, relative to an arbitrary line in the plane of motion, must satisfy

$$\theta_2(t_d) = \theta_1(t_d) + (\pi - n_2 T)$$

where $t_d$ is the time of departure, T is the transfer time and $n_2$ is the mean motion of the target. For Hohmann transfers to an outer planet, the last term is positive and the target must “lead” the departure planet. Conversely for transfers to an inner planet. For transfers from the Earth to Mars the lead angle is about 43 deg. and from equatorial LEO to GEO is about 100 deg. By eliminating T, a general expression for lead angle can be derived in terms of only the semi-major axis ratio. Since the relative angular motion is $|n_2 - n_1|$, the relative configuration will reoccur every **synodic period** given by $\frac{2\pi}{|n_2 - n_1|} = \frac{P_1 P_2}{|P_1 - P_2|}$, where P represents orbital period. For the Earth and Mars the synodic period is about 26 months, thus near minimal energy launch opportunities occur every 26 months. For the equatorial LEO to GEO, the synodic period is about 100 minutes. This difference places substantially more pressure on interplanetary launches than on most launches into near Earth orbits.

Exercise 6-5. Calculate the synodic period and lead angle for 180 deg. transfers from the Earth to each of the planets. Ignore planetary eccentricities and inclinations. Plots results vs. planetary semi-major axis in AU and interpret results.

### 6.3.4 Bi-elliptic and bi-parabolic transfer

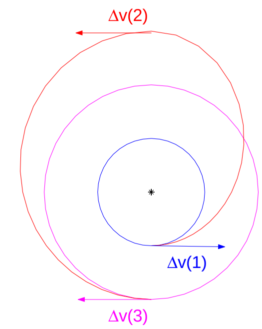



Figure 6-4. Bi-elliptic transfer.

It has been shown that the Hohmann transfer is not the global minimum energy transfer if the ratio of major axis is greater than 11.94. Above this ratio the **bi-parabolic** transfer is the optimal. In this transfer, $\Delta v_1$ is performed with $\gamma_1$=0 to establish a parabolic orbit. At $r = \infty$ the second maneuver is performed to adjust the angular momentum so that the return parabola will have a radius of periapsis of $r_2$. Since the maneuver is at infinity, these changes can be made with $\Delta v_2$=0. The third maneuver on the return parabola circularizes the orbit at $r_2$. Such a transfer is of course impractical. For large semi-major axis ratios, **bi-elliptic** transfers can provide a small reduction it total Δv, but at a significant increase in transfer time. A bi-elliptic transfer is shown in Figure 6-4. LEO to GEO transfers can not benefit from the bi-elliptic option and only transfers from Earth to Uranus, Neptune and Pluto could benefit. The second maneuver must occur outside the outer orbit, so the transfer time is more than doubled for a small saving in Δv [1,111]. Thus for interplanetary transfers this option is also impractical.

### 6.3.5 Impulsive Transfers Between Inclined Orbits

Transfers of this type include the transfer from an inclined LEO to the GEO and the interplanetary transfer between Earth and any planet. In the first case the difference in inclination can be 28 degrees or more, while in the second case the inclination difference is at most a few degrees. From equations (5-27) an impulse is most effective for making small changes in inclination when $r\cos\theta$ achieves its maximum value. For circular orbits this is at the ascending or descending node. For eccentric orbits the most effective location depends on the particular values of eccentricity and $\omega$ and corresponds to the true anomaly that satisfies

$$\tan(f + \omega) = \frac{e\sin f}{1 + e\cos f} \tag{6-19}$$

This equation can be reduced to $\sin(f+\omega)+e\sin\omega=0$, from which two solutions for extrema can be found.

For large angle plane changes the assumptions associated with equations (5-27) are not applicable and finite velocity changes must be included. Consider first the case where the orbits are circular and have the same semi-major axis, a. This is the case of rotating the initial orbit to a new inclination. The transfer can be accomplished with a single impulsive maneuver of magnitude $\Delta v = 2\sqrt{\frac{\mu}{a}}\sin\frac{\Delta i}{2}$, where $\Delta i$ is the difference in inclination between the two orbits. With the bi-parabolic generalization of the Hohmann transfer as a guideline, consideration might be given to a three maneuver bi-parabolic transfer for inclination changes. The first maneuver places the vehicle on a parabolic trajectory without changing the orbit plane $\Delta v_1 = \sqrt{\mu/a}(\sqrt{2}-1)$, the second maneuver is performed at infinity to change the orbit inclination to the desired value without changing the periapsis distance $\Delta v_2=0$, and the third maneuver changes the return parabolic motion to the original circular orbit $\Delta v_3=\Delta v_1$. It is readily shown that the bi-parabolic plane change is less expensive than the single impulse plane change if $\Delta i>48.9^o$. Continuing the analogy with the co-planar circle to circle transfers, there are three impulse bi-elliptic maneuvers that require less $\Delta v$ than single impulse plane change maneuvers if $\Delta i>38.9^o$ [4,118]. All three maneuvers are performed on the line of nodes. This approach of course extends the time to achieve the final orbit. Further savings in $\Delta v$ can be achieved by performing a fraction of the ultimate plane change at each of the three maneuvers. However, plane changes of this magnitude are generally avoided because of the large propulsion penalty.

Now consider the case of transfer between inclined circular orbits of radius $a_1$ and $a_2$. The bi-parabolic and bi-elliptic transfers again provide advantages in particular cases, but emphasis here will be on the two impulse transfer at the line of nodes. First order analysis would suggest that the plane changed be performed at minimum velocity, i.e. at the larger orbit. However there is an advantage to performing part of the plane change during both maneuvers. The magnitude of the velocities before and after each maneuvers are the same as for the Hohmann transfer. The $\Delta v$'s are different because the velocities are not co-linear. The velocity before the first burn is the circular speed from the first orbit:

$$v_1^- = \sqrt{\frac{\mu}{a_1}} \; . \tag{6-20}$$

The speed on the transfer orbit immediately after the first burn is

$$v_1^+ = \sqrt{2\mu\left(\frac{1}{a_1} - \frac{1}{a_1 + a_2}\right)} \tag{6-21}$$

and before the second impulse

$$v_2^- = \sqrt{2\mu\left(\frac{1}{a_2} - \frac{1}{a_1 + a_2}\right)} \tag{6-22}$$

On the final orbit, after the second impulse

$$v_2^+ = \sqrt{\frac{\mu}{a_2}} \tag{6-23}$$

Let the total required plane change be divided between the two maneuvers $\Delta i = \Delta i_1+\Delta i_2$. From the law of cosines, the speed change required to transition to the transfer orbit is

$$\Delta v_1 = \sqrt{{v_1^+}^2 + {v_1^-}^2 - 2v_1^+ v_1^- \cos\Delta i_1} \tag{6-24}$$

which is a function of only $\Delta i_1$. Likewise, the second impulse

$$\Delta v_2 = \sqrt{{v_2^+}^2 + {v_2^-}^2 - 2v_2^+ v_2^- \cos(\Delta i - \Delta i_1)} \tag{6-25}$$

is also a function of only $\Delta i_1$ since all the other parameters are specified. The total impulsive requirement can be directly minimized. For the transfer from LEO to GEO with a $28^o$ plane change, the minimum $\Delta v$ occurs if about $2^o$ of plane change is performed at departure and $26^o$ at arrival.

### 6.3.6 Other Impulsive Transfers

The results of the above sections can be used as starting points for numerical searches for the case of two or three impulse transfers between inclined circular orbits. Solution approaches range from multi-dimensional optimization to zeroth order searches of the design space. Two impulse transfers between slightly elliptic, nearly co-planar orbits can also be optimized by constraining the solution to transfer angles near $180^o$. A particularly useful form for calculating velocity vectors before or after an impulsive maneuver can be obtain by crossing the angular momentum vector with equation (3-6) to get

$$\mathbf{v} = \mathbf{h} \times (\mathbf{e} + \mathbf{e_r}) / p \qquad (6\text{-}26)$$

which provides velocity as a function of p, e, i, $\Omega$, $\omega$ and f. The true anomaly enters the equation through the unit vector along the radial direction, $\mathbf{e}_r$, thereby providing a convenient method for calculating velocity around the orbit as a function of true anomaly.

## 6.4 Low Thrust Transfer

The discrete impulsive maneuvers discussed above would normally be performed with rocket engines having specific impulses less than 500 sec. A potential alternative is a transfer using continuous, low thrust, high specific impulse technology. Ion propulsion systems provide specific impulses that are up to 10 times those of chemical systems. However, they are not applicable for missions requiring high accelerations or short mission times. Orbit transfer optimization associated with continuous thrust trajectories is difficult and requires the application of special numerical techniques. There are three special cases, tangential, circumferential, or radial acceleration, where approximate analytic solutions can be formed [3]. This section will present a numerical example for constant tangential acceleration and outline an analytic solution for the constant circumferential acceleration case.

### 6.4.1 Constant Tangential Acceleration Escape Trajectories

Assume the spacecraft is initially in a circular orbit with semi-major axis $a_o$. A constant tangential acceleration is applied until the s/c reaches escape velocity. Since the thrust is applied in the same direction as the velocity vector, the thrust is providing the maximum change in kinetic energy at every instant of time. With the acceleration $a_T$ applied tangent to the orbit, the equations of motion are

$$\frac{d^2x}{dt^2} + \frac{\mu x}{r^3} = a_T\left(\frac{\dot{x}}{v}\right) \qquad \frac{d^2y}{dt^2} + \frac{\mu y}{r^3} = a_T\left(\frac{\dot{y}}{v}\right) \qquad (6\text{-}27)$$

where $a_T$ is the magnitude of the acceleration. Numerical integration of these equations results in the spiral trajectory shown in Figure 6-5. The assumption of constant acceleration implies a variable thrust level for a rocket system. Solar sail systems would not be expelling mass, but generally cannot provide constant tangential thrust because the distance from the Sun and the orientation to the Sun change throughout the orbit. At the acceleration level of $0.005g_o$, the spacecraft reaches escape velocity (the red star) after about 33 hours and 7 revolutions of the Earth. For comparison, if the acceleration is reduced to $0.001g_o$, it will take about 180 hours to reach escape velocity.

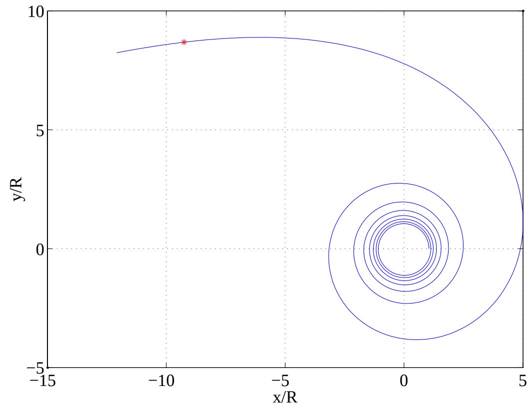



Figure 6-5. Escape from LEO (h=300 km) using constant tangential thrust of $0.005g_o$

## 6.4.2 Constant Circumferential Acceleration

Battin [3,418] provides the following as an exercise for the student. The constant circumferential acceleration $a_\theta$ is directed perpendicular to the radius vector, and the equations of motion are

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{\mu}{r^2}$$
$$\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = ra_\theta \tag{6-28}$$

If the radial acceleration is small so that the centripetal acceleration balances gravity, the equation for distance from the center of attraction is

$$r = \frac{r_o}{\left[1 - \frac{(t-t_o)a_\theta}{v_o}\right]^2} \tag{6-29}$$

where $v_o$ is the velocity in the initial circular orbit. The radius at escape is

$$r_{esc} = \frac{r_o v_o}{(2a_\theta r_o)^{1/2}} \tag{6-30}$$

and the time to escape is

$$t_{esc} = t_o + \frac{v_o}{a_\theta}\left[1 - \left(\frac{2a_\theta r_o}{v_o^2}\right)^{1/4}\right] \tag{6-31}$$

The trajectory is again a spiral as in Figure 6-5. As might be expected, the tangential thrust reaches escape velocity sooner than the circumferential thrust case.

## 6.5 Lambert's Theorem

For rendezvous and interception type orbit transfers, the orbital elements of both the departure and target orbits are known, i.e. the position in both orbits are known functions of time. The transfer orbit must therefore pass through a point on the departure orbit at the departure time and pass through a point on the target orbit at a later time, the arrival time. Thus, six state variables and a transfer time are specified for each possible transfer. It is natural to ask if more than one orbit can satisfy these conditions. The answer is partially given by Lambert's theorem of 1761 which states that the time required to transfer between two points $P_1$ and $P_2$ on an elliptic transfer orbit depends only on the semi major axis of the ellipse, the chord length c and the sum of the radii from the focus to point $P_1$ and $P_2$.

$$\sqrt{\mu}(t_2 - t_1) = f(a, r_1 + r_2, c) \tag{6-32}$$

Lambert's theorem involves conditions at two times and is referred to as a **boundary value problem**. The most extensive coverage of the topic is given in [3,237]. Since Kepler's equation only involves one time, its solution is called an **initial value problem**.

### 6.5.1 Proof of Lambert's theorem

The first analytic proof of Lambert's theorem was given by Lagrange in 1778. The proof develops a form of Kepler's equation that is only a function of the desired variables. Let $E_2$ and $E_1$ denote the eccentric anomalies at time $t_2$ and $t_1$, respectively, then Kepler's equation (3-14) can be written as

$$\sqrt{\mu}(t_2 - t_1) = a^{\frac{3}{2}}[E_2 - E_1 - e(\sin E_2 - \sin E_1)] \tag{6-33}$$

Since $\sin E_2 - \sin E_1 = 2\sin\frac{E_2 - E_1}{2}\cos\frac{E_2 + E_1}{2}$, the transfer time can be written as the sum and difference of the eccentric anomalies. So it is convenient to define

$$E_p = \frac{1}{2}(E_2 + E_1)$$
$$\text{and} \tag{6-34}$$
$$E_m = \frac{1}{2}(E_2 - E_1)$$

Likewise, since r=a(1-e cos E), the sum of $r_1$ and $r_2$ can be written as

$$r_1 + r_2 = a[2 - e(\cos E_1 + \cos E_2)] = 2a(1 - e\cos E_p \cos E_m) \tag{6-35}$$

To calculate the chord length from $P_1$ to $P_2$ use the orbital coordinate system from Section 3.8:

$$\xi = a\cos E \qquad \eta = b\sin E \qquad b = a\sqrt{1 - e^2} \tag{6-36}$$

The chord distance in terms of a and e is then

$$\begin{aligned} c^2 &= (\xi_2 - \xi_1)^2 + (\eta_2 - \eta_1)^2 \\ &= a^2[(\cos E_2 - \cos E_1)^2 + (1 - e^2)(\sin E_2 - \sin E_1)^2] \\ &= 4a^2 \sin^2 E_m (1 - e^2 \cos^2 E_p) \end{aligned} \tag{6-37}$$

Since e is less than 1 for elliptic orbits, define

$$\cos\kappa = e\cos E_p \tag{6-38}$$

A quadrant ambiguity has been introduce here which would have to be resolved in practice. Nevertheless, equation (6-37) then reduces to

$$c = 2a\sin E_m \sin\kappa > 0 \tag{6-39}$$

and equation (6-35) becomes

$$r_1 + r_2 = 2a(1 - \cos E_m \cos\kappa) \tag{6-40}$$

The transfer time given by equation (6-33) in terms of these newly defined parameters is

$$\sqrt{\mu}(t_2 - t_1) = 2a^{\frac{3}{2}}[E_m - \cos\kappa \sin E_m] \tag{6-41}$$

Ignoring the quadrant issues, equations (6-39) and (6-40) can be used to solve for $E_m$ and $\kappa$ given only a, $r_1$+$r_2$ and c. Thus Lambert's theorem is proved and as stated the transfer time only depends on $r_1$+ $r_2$, c, and a.

### 6.5.2 Euler's equation for parabolic orbits

A special case of Lambert's theorem which describes the parabolic transfer time between two points, was derived by Euler in 1743. The result is [3, 277]

$$\sqrt{\mu}(t_2 - t_1) = \frac{1}{6}\left[(r_1 + r_2 + c)^{\frac{3}{2}} \pm (r_1 + r_2 - c)^{\frac{3}{2}}\right] \tag{6-42}$$

with the negative sign if $\theta=f_2-f_1 < \pi$, and positive for $\theta > \pi$. During searches for feasible departure and arrival times, this equation is useful for identifying when the transfer switches from elliptic to parabolic.

## 6.6 Interception and Rendezvous

In the standard approach to determining interception and rendezvous transfer orbits, a set of potential departure times and a set of potential arrival times are selected. For interplanetary and other transfers requiring small inclination changes, the sets are generally centered around the Hohmann or 180$^o$ transfer opportunities. For transfers that require large plane changes a broader search space may be required. For each departure time, an initial position vector $\mathbf{r}_1$ is determined.

Likewise for each arrival time, $\mathbf{r}_2$. Lambert's theorem solutions for the transfer orbit are performed for every pair of departure dates and arrival dates. Contour plots of mission parameters, e.g. $\Delta v_1$, $\Delta v_2$ are then plotted as functions of departure and arrival date. In this approach, the only unknown in equation (6-41) is the semi major axis, a. Thus iterative schemes must be utilized to find the semi major axis which satisfies the equation. Many methods have been developed [1,3,6] to obtain a solution. Even today no method is accepted as the "best" and many methods fail for the $180^o$ transfer. The method presented here is by Herrick and Liu as described in Reference 7 and is one form of the class of solutions called "p-iteration."

The procedure follows the steps:

1. Determine the transfer angle $0<\theta=f_2-f_1<2\pi$ from the dot product $\mathbf{r_1} \cdot \mathbf{r_2}$ and the cross product $\mathbf{r_1} \times \mathbf{r_2}$. For transfers greater than $180^o$, care must be exercised to assure angular momentum of the transfer orbit is in the desired direction. Without such care, retrograde transfer orbits may result from transferring between two direct orbits.
2. Pick an initial estimate of the semi-latus rectum of the transfer, p.
3. Solve for the F and G functions from Section 3.8

$$F(f_1,f_2) = 1 + \frac{r_2}{p}[\cos(f_2 - f_1) - 1] \qquad G(f_1,f_2) = \frac{r_2 r_1}{\sqrt{\mu p}}\sin(f_2 - f_1)$$

4. Determine the initial velocity that will produce such a transfer: $\mathbf{v_1} = \frac{1}{G}(\mathbf{r_2} - F\mathbf{r_1})$. Note that 0 and 180 degree transfers are not permitted. The resulting transfer time may not be the desired value since only the difference in true anomaly has been use.
5. Using $\mathbf{r}_1$, $\mathbf{v}_1$ and $t_1$ map the estimated transfer orbit to time $t_2$ to get the position vector $\tilde{\mathbf{r}}$.
6. Iterate on p until the angle between $\tilde{\mathbf{r}}$ and $\mathbf{r}_1$ is $\theta$.

By starting near the 180 degree transfer, an initial guess for p is readily available. A Newton-Raphson iteration scheme can be utilized if finite difference partials of $\mathbf{r_1} \cdot \tilde{\mathbf{r}}$ wrt p are generated.

## 6.7 Midcourse Corrections

**Midcourse maneuvers** or **trajectory correction maneuvers** are performed throughout the life of most space missions. The need for such maneuvers generally result from three causes:

1. **Errors in modeling the forces acting on the satellite**. There are always uncertainties in the modeling of forces acting on a satellite. Precise calculation of radiation pressure is difficult because the optical properties of materials are not known at all angles of incidence and wavelengths. Further, reflected and emitted radiation from the Earth is very dependent on geographic location and cloud cover. Gravitational terms are in error because the masses of the planets, Moon and Sun and the non-central gravity field are uncertain. Drag and other aerodynamic forces are difficult to model to better than 10% accuracy because of uncertainties in both the aerodynamic properties of the spacecraft and the atmospheric environment. These

types of errors usually produce small but often secular deviations of the trajectory from the nominal path and are corrected whenever the orbit approaches some mission tolerance limit.

2. **Execution errors in previous propulsive maneuvers.** Maneuver errors result from errors in pointing the spacecraft to perform the maneuver and errors in the total Δv imparted. The former depend on both the knowledge of spacecraft orientation at the beginning of the burn and the control of the orientation during the burn. The latter errors can be due to a number of causes including uncertainties in the thrust level for the specific maneuver, burn time, and blown-down after the fuel supply is terminated. Maneuvers are monitored very closely to assure that the performance was not outside the nominal error limits. If an off nominal maneuver was performed, perhaps due to a sticky propellant valve, the need for an immediate corrective maneuver is determined. Designing the time between nominal maneuvers and the Δv allocation are part of the mission and operations designs.

3. **Changes in mission requirements.** Mission requirements can also change during the course of a mission. If "large" orbit changes are required to accomplish the new mission, the maneuver methods discussed in sections (6.3), (6.4) and (6.6) are appropriate. Many of these methods require iterative or other multi-step numerical methods. If the maneuvers are "small" then linear approximation methods can be used to determine the maneuver or at least obtain a very good estimate of the maneuver which can be verified or slightly adjusted through precise trajectory calculations.

### 6.7.1 State transition matrix

The fundamental tool for performing small, impulsive maneuver design and analysis is the **state transition matrix**. The state transition matrix is obtained by linearizing the equations of motion, equation (2-1) along a **reference trajectory**. The "two body" state transition matrix is obtained if the only force is due to the central gravity term. The general state transition matrix is obtained first and then specialized to the two body case. Write Newton's second law in the six-vector form of equation (5-8)

$$\frac{d\mathbf{r}}{dt} = \mathbf{v} \qquad \frac{d\mathbf{v}}{dt} = \mathbf{f}(\mathbf{r},\mathbf{v},t) \tag{6-43}$$

Assume that initial conditions $\mathbf{r}_o$ and $\mathbf{v}_o$ are given at time $t_o$ and the resulting solution is the reference or **nominal trajectory** $\mathbf{r}(\mathbf{r}_o,\mathbf{v}_o t)$ and $\mathbf{v}(\mathbf{r}_o,\mathbf{v}_o t)$. If the initial conditions are perturbed by some small amount $\delta\mathbf{r}_o$ and $\delta\mathbf{v}_o$ the change in position and velocity at time t would be

$$\delta\mathbf{r}(t) = \frac{\partial\mathbf{r}(t)}{\partial\mathbf{r}_o}\delta\mathbf{r}_o + \frac{\partial\mathbf{r}(t)}{\partial\mathbf{v}_o}\delta\mathbf{v}_o \qquad \delta\mathbf{v}(t) = \frac{\partial\mathbf{v}(t)}{\partial\mathbf{r}_o}\delta\mathbf{r}_o + \frac{\partial\mathbf{v}(t)}{\partial\mathbf{v}_o}\delta\mathbf{v}_o \tag{6-44}$$

where only the linear terms in the Taylor series expansion have been retained and the partial derivatives are evaluated along the nominal trajectory. These equations can be combined in the form

$$\begin{bmatrix}\delta\mathbf{r}\\ \delta\mathbf{v}\end{bmatrix}_t = \begin{bmatrix}\mathbf{R}_r & \mathbf{R}_v\\ \mathbf{V}_r & \mathbf{V}_v\end{bmatrix}\begin{bmatrix}\delta\mathbf{r}\\ \delta\mathbf{v}\end{bmatrix}_o = \Phi(t,t_o)\begin{bmatrix}\delta\mathbf{r}\\ \delta\mathbf{v}\end{bmatrix}_o \tag{6-45}$$

where $\mathbf{R}_r$, $\mathbf{R}_v$, $\mathbf{V}_r$ and $\mathbf{V}_v$ are the 3 by 3 submatrices from equation (6-44) and the **state transition matrix** is defined as:

$$\Phi(t, t_o) = \begin{bmatrix} \frac{\partial x(t)}{\partial x(t_o)} & \frac{\partial x(t)}{\partial y(t_o)} & \cdots & \cdots & \cdots & \frac{\partial x(t)}{\partial \dot{z}(t_o)} \\ \frac{\partial y(t)}{\partial x(t_o)} & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \dot{z}(t)}{\partial x(t_o)} & \cdots & \cdots & \cdots & \cdots & \frac{\partial \dot{z}(t)}{\partial \dot{z}(t_o)} \end{bmatrix} \tag{6-46}$$

The terms in this equation are obtained by integrating the equations of motion resulting from differentiating equations (6-43) with respect to each of the 6 initial conditions; for example, for the initial position component $x_o$

$$\frac{d}{dt}\left(\frac{\partial \mathbf{r}}{\partial x_o}\right) = \left(\frac{\partial \mathbf{v}}{\partial x_o}\right) \qquad \frac{d}{dt}\left(\frac{\partial \mathbf{v}}{\partial x_o}\right) = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{r}}\right]\left(\frac{\partial \mathbf{r}}{\partial x_o}\right) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{v}}\right]\left(\frac{\partial \mathbf{v}}{\partial x_o}\right) \tag{6-47}$$

This is a set of 6 coupled, first order, linear, non-autonomous, homogeneous, differential equations for the six quantities in brackets ( ). The terms in braces [ ] are 3 by 3 matrices that are evaluated along the nominal trajectory and therefore known functions of time. Note that there is no coupling between equations for different components of the initial conditions. Secondly, if the force does not depend on velocity, the last term in the second equation vanishes. The equations for all six of the initial conditions can be written as

$$\frac{d}{dt}\Phi = \dot{\Phi} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{I}_3 \\ \frac{\partial \mathbf{f}}{\partial \mathbf{r}} & \frac{\partial \mathbf{f}}{\partial \mathbf{v}} \end{bmatrix}\Phi \tag{6-48}$$

with initial conditions $\Phi(t_o,t_o) = \mathbf{I}_6$. This is the equation of motion for the **state transition matrix**. For two body motion $\mathbf{f}=-\mu\mathbf{r}/r^3$ so that $\partial\mathbf{f}/\partial\mathbf{v} = \mathbf{0}$ and

$$\frac{\partial \mathbf{f}}{\partial \mathbf{r}} = \frac{-\mu}{r^3}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{3}{r^2}\begin{bmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{bmatrix}\right) \tag{6-49}$$

Explicit expressions for the elements of the two body state transition matrix can be derived after considerable algebra, one representation in given in (3,467). For the general case, numerical integration is the only resort and many orbit propagation programs will simultaneously integrate equations (6-43), equations (6-48) and generalizations of (6-48).

The **distinction** between the state transition matrix and the F and G functions of Section 3.8 is that the F and G functions are used to map the reference 2-body trajectory from one time to another while the state transition matrix is used to map small deviations from the reference trajectory from one time to another.

### 6.7.2 Constant time of arrival maneuvers

Consider a spacecraft that has been targeted to arrive at a particular point in space $\mathbf{r}_t$ at some specified time, $t_t$ on the nominal trajectory. Examples include interception and interplanetary missions. Because of various errors mentioned above, the trajectory is not going to meet the objective. Orbit determination and trajectory calculations predicts that at time $t_t$ on the current trajectory, the spacecraft will be at position $\mathbf{r}_s$. The error $\delta\mathbf{r}=\mathbf{r}_t-\mathbf{r}_s$ is to be corrected at time $t_m<t_t$ with an impulsive $\delta\mathbf{v}$ maneuver. Evaluating the state transition matrix $\Phi(t_t,t_m)$ along the current trajectory provides from equation (6-45)

$$\delta\mathbf{r}(t_t) = \mathbf{R}_v(t_t,t_m)\delta\mathbf{v}(t_m) \tag{6-50}$$

If the inverse of $\mathbf{R}_v$ exist, the required impulsive maneuver at time $t_m$ is readily calculated. Since the impulse does not change the position, $\delta\mathbf{r}(t_m)=0$ and this term in equation (6-45) does not contribute to either the final position or velocity. The equation shows that velocity at time $t_a$ is influenced by the maneuver, but this term is often ineligible compared to the nominal velocity at $t_t$. If final velocity is a consideration, then a second maneuver could be performed at time $t_t$ to correct for any velocity mismatch. The are orbital transfers for which $\mathbf{R}_v$ is singular; for example, an impulse at any point in an orbit can not change the position normal to the orbit plane for a $180^o$ transfer. In this case $\mathbf{R}_v$ would be at most of rank two.

In this problem there are no extra degrees of freedom, i.e. there are three positions to be corrected and three components of velocity to be changed. In some cases, the target may be a two dimensional surface or a line in three dimensional position space. For example, suppose for a rendezvous mission a sequence of N trajectory correction maneuver are required to meet the final accuracy requirements. More than one maneuver is required because early orbit determination and maneuver execution errors may be large. To avoid possible collision if subsequent maneuvers fail, only the N-th maneuver can target for rendezvous. The first N-1 maneuvers are then targeted to sequentially closer distances which might be represented by ellipsoids surrounding the target. In this case there is one or two extra degrees of freedom in choosing the impulsive maneuver and minimum $\Delta$v maneuvers become a consideration.

### 6.7.3 Variable time of arrival maneuvers

For both interception and interplanetary missions, small variations in arrival time may be acceptable. Two types of problems can be now posed: (1) for a fixed $\Delta\mathbf{v}$ magnitude determine the earliest time of arrival and (2) for a variable time of arrival determine the minimum $\Delta\mathbf{v}$. Within the linear approximations of this section, both formulations are very similar, so consider the latter case. At time $t_t$, let $\mathbf{r}_s$ and $\mathbf{v}_s$ be the spacecraft position and velocity and $\mathbf{r}_t$ and $\mathbf{v}_t$ be the target position, then within the linear assumptions

$$\begin{aligned} \mathbf{r}_s(t_t + \delta t) &= \mathbf{r}_s(t_t) + \mathbf{v}_s(t_t)\delta t + \mathbf{R}_v(t_t, t_m)\delta\mathbf{v}(t_m) \\ \mathbf{r}_t(t_t + \delta t) &= \mathbf{r}_t(t_t) + \mathbf{v}_t(t_t)\delta t \end{aligned} \tag{6-51}$$

where the $\mathbf{R}\delta\mathbf{v}$ term represents the change in spacecraft position at time $t_t$ due to an impulsive change in velocity at the maneuver time $t_m$. Small changes in $\mathbf{v}_s$ due to the maneuver are ignored as second order. The goal is to make $\mathbf{r}_s(t_t+\delta t)=\mathbf{r}_t(t_t+\delta t)$ by selecting $\delta t$ so as to minimize $\delta\mathbf{v}$. Setting these terms equal leads to the solution

$$\delta\mathbf{v} = \mathbf{R}_v^{-1}[\delta\mathbf{r} + \nu\delta t] \equiv \mathbf{a} + \mathbf{b}\delta t \tag{6-52}$$

where $\delta\mathbf{r}=\mathbf{r}_t-\mathbf{r}_s$ as above and $\nu=\mathbf{v}_t-\mathbf{v}_s$ is the relative velocity. It is straight forward to show that the time that minimizes the norm of $\delta\mathbf{v}$ occurs when $\delta t = -\frac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}}$. The minimum impulse to provide interception is therefore

$$\delta\mathbf{v} = \mathbf{a} - \left(\frac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}}\right)\mathbf{b} \tag{6-53}$$

Exercise 6-6. Starting with equation (6-52), verify and provide a physical interpretation for equation (6-53).

## 6.8 Problems

1. Consider the vertical launch of a two stage vehicle with a 1000 kg payload and a first stage structural plus propellant mass of 100,000 kg, $\beta$=0.04, $I_{sp}$=300s, and thrust of 1.5e6 N. The respective second stage values are 5,000, 0.08, 350, and 180,000. If gravity and drag are neglected, what is the burn out velocity? Is this enough to escape the Earth?
2. For the same rocket as in problem 1, assume a uniform gravity field of 10 m/s$^2$ and no other external forces. What is the burn out velocity? Is this enough to escape the Earth?
3. A Russian satellite is in a circular LEO at 400 km altitude and inclination 55 deg. Assume a two impulse transfer to a geosynchronous circular equatorial orbit. How should the inclination change be divided among the two impulses to minimize the total $\Delta V$?
4. Consider 3-d particle motion in a uniform gravity field ($g_o$) along the z axis and initial conditions ($x_o$, $y_o$, $z_o$, $u_o$, $v_o$, $w_o$) at time $t_o$. Write the state transition matrices $\Phi(t,t_o)$ and $\Phi(t,t_m)$ where $t_o<t_m<t$ is a potential midcourse maneuver time.

## 6.9 References


1. Prussing, J.E., and Conway, B.A. *Orbital Mechanics*, Oxford University Press, New York, 1993.
2. Griffin, M., and French, J., *Space Vehicle Design*, AIAA Educational Series, New York, 1991.

3. Battin, R.H., *An Introduction to the Mathematics and Methods of Astrodynamics*, AIAA Educational Series, New York, 1987.
4. Chobotov, V.A., *Orbital Mechanics*, AIAA Educational Series, Washington, 1991.
5. Lawden, D.F., *Optimal Trajectories for Space Navigation*, Butterworth & Co., London, 1963.
6. Bate, R.R., Mueller, D.D., and White, J.E., *Fundamentals of Astrodynamics*, Dover Publications, Inc. New York, 1971.
7. Danby, J.M.A., *Fundamentals of Celestial Mechanics*, Willmann-Bell, Richmond, Va., 1989.

# Chapter 7 - Special Orbits

## 7.1 Introduction

An important consideration in space mission design is determining the type of Earth orbit that best suits the design goals and purpose of the mission. This chapter will include a discussion of the advantages and disadvantages of these orbits with emphasis on the geosynchronous, Sun-synchronous, Molniya, polar, low Earth orbits and frozen orbits. There will also be a discussion of the ground track characteristics of certain orbits and the possible benefits of the use of constellations.

## 7.2 Geosynchronous Orbits (GEO)

Geosynchronous orbits are a special class of **synchronous orbits**. A synchronous orbit will have a repetitive ground track after an integer number of days (m) and an integer number of orbits (n). Usually m is a small integer but may be 30 days or more. The value of n of course depends on the period of the orbit. The smallest practical value of n is 16 and even this orbit may have a relatively short life time due to drag. If m/n<1 the orbit is **sub-synchronous** (e.g. for a 8 hour orbital period m/n=1/3) and if m/n>1 the orbit is **super-synchronous** (e.g. for a 36 hour orbit m/n=3/2).

Of all the synchronous orbits, the geosynchronous orbit is the most utilized type, with many of the meteorological and communication satellites in this orbit. A **geosynchronous** orbit has an orbital period equal to one sidereal day (1.3.5) so m=n=1. The benefit of the geosynchronous orbit is that a satellite in this orbit will appear over the same point on the Earth at the same time each day. The statement is only precise within the framework of the 2-body problem since perturbations (Chapter 5) will produces deviations and must be considered in the design of the orbit. The special case of a circular geosynchronous orbit that has zero inclination to the equator is referred to as **geostationary**. A satellite in a geostationary orbit will appear to be motionless to an observer on the surface of the Earth with no change in elevation or azimuth during the day. This orbit is useful for communication satellites because it eliminates the need for tracking mechanisms so that communication with the satellite can be maintained with a fixed antenna at the Earth tracking station. The commercial importance of this orbit is demonstrated by an international agreement that assigns positions in the geostationary orbit. Original spacing was limited to 3 degrees in longitude but this was subsequently reduced to 2 degrees. The spacing is limited by radio interference criteria and not orbital considerations.

The sidereal day (Section 1.3.8) is approximately $23^h56^m04^s$ or 86,164 seconds. From Section 3.3.5 the two body semi-major axis for a geosynchronous satellite would give a=42,164 km. Additional benefits of this high altitude include negligible atmospheric drag, being above the Van Allen radiation belts, and providing coverage of the Earth's surface up to about $82^o$ latitude. The main economic drawback is the launch cost. Establishing a geostationary orbit generally consists of three phases. First, the satellite is inserted into a low Earth orbit (LEO). This orbit provides an opportunity to check spacecraft health and establish proper phasing for the orbit transfer. Next a near Hohmann transfer from the LEO to GEO is executed. This geosynchronous transfer orbit (GTO) has a periapsis near the LEO and the apoapsis near the GEO. Finally a maneuver is performed near apoapsis on the GTO to circularize the orbit and to make the inclination near zero. A plane change of up to $28.5°$ is needed for a satellite launched from Kennedy Space Center. For this case, the typical $\Delta V$ for transfer from LEO to GEO is greater than 4.2 km/s, whereas from LEO the $\Delta V$ to escape the Earth is only 3.2 km/s. Establishing a GEO from an equatorial launch site (zero inclination LEO) requires about 3.9 km/s total $\Delta V$.

GEO satellites must be maintained within their assigned longitudinal spaces. This **station keeping** requirement can produce a significant propulsive penalty. The major perturbations of a GEO satellite are $J_2$ and to a much lesser extent other non-central gravity terms (Section 5.4.2), solar pressure (Section 5.4.5), and N-body perturbations from the Moon and Sun (Section 5.5.2). Even at this altitude, $J_2$ still produces the largest precession in the orbit periapsis, but solar pressure and lunar and solar gravity produce long period changes in eccentricity and/or inclination. These later effects can cause the satellite to move out of the assigned region. N-body effects can produce long period variations in inclination greater than $15^o$ over times scales of a decade. Orbits are typically established with a small inclination, so that after a 3 to 5 years the perturbations drive the inclination to zero and after 3 to 5 more years the inclination has again increased to a few degrees. Proper design can assure inclinations less than $5^o$ over the life of the spacecraft. At the end of the useful life, geostationary satellites are generally moved to higher altitude orbits and powered off so as not to interfere with GEO or GTO orbits.

Exercise 7-1. Assume a GEO has been assigned a location of $75^o$ W longitude. Make plots of longitude vs. time, latitude vs. time, and latitude vs. longitude for inclination (deg.) and eccentricity pairs of (4,0), (10,0), (0,0.002), (0,0.004), and (4,0.004). Multi-plots per page are recommended but, to provide easy comparison, plots of the same type must have the same axis range. Provide all calculations, m-files and a discussion of the results considering at least fixed Earth antenna pointing and longitudinal station keeping.

## 7.3 Molniya Orbits

The Molniya orbit is a **sub-synchronous** orbit that has a period of 12 hours so m=1 and n=2. The satellite will thus pass over the same point every other orbit. Note that there is a lack of precision in the synchronous orbit definitions and that discipline jargon is often used. By a "12 hour orbit" it is recognized that this means the ground track repeats after two orbits considering significant orbital perturbations and the sidereal rotation of the Earth. Since 1965 over 100 Molniya satellites have been launched. The Molniya orbit is highly inclined to provide radio communication across wide areas of the Soviet Union. The orbit is also highly elliptical to reduce launch requirements. This combination permits a Molniya spacecraft to remain over high latitudes twice each day for extended periods of time. The high eccentricity means that a satellite only spends about 3 hours over the southern hemisphere. To maintain the apoapsis at a high northern latitude, the inclination is selected near the critical inclination (Section 5.4.3) of $63.4^o$ or $116.6^o$ so that $\dot{\omega} = 0$. Of course, the argument of perigee does not remain constant because of higher order gravity terms and the Sun-moon attraction. These effects are included in the design of the orbit and remaining perturbations are adjusted by orbit maintenance maneuvers. The typical Molniya orbit has a period of about one-half of a sidereal day, a semi-major axis of about 26,500 km, and the eccentricity is between 0.72 and 0.75.

Exercise 7-2. Plot a one day ground track for both a direct and a retrograde Molniya satellite. Consider only $J_2$ perturbations and include $J_2$ precession in calculating the desired orbital period. Pick a periapsis altitude between 400 and 1000 km, apoapsis at maximum northern latitude, and longitude of apoapsis of 90 deg. east. Show tick marks along the ground track every 30 minutes. Use the supplied m-file "GroundTrack.m" to put the ground track on the globe. Also plot altitude vs. latitude over one orbital period with the same tick marks. Show all calculations, m-files, and discuss results including advantages and disadvantages of the direct and retrograde cases.

LEO parking orbits near the critical inclination are easily obtainable from the high latitude launch sites in Russia. The $\Delta V$ required to place a satellite into a Molniya orbit from a critically inclined parking orbit is about 2.5 km/s. This is significantly lower than the delta V needed to place an satellite into geosynchronous orbit. The disadvantage is of course that ground tracking stations must include some sort of antenna pointing control.

## 7.4 Polar Orbits

For full global coverage of the Earth, a ground track would have to cover latitudes up to $\pm 90°$. The only orbit that satisfies this condition has an inclination of 90°. These types of orbits are referred to as **polar orbits**. Polar orbits are used extensively for the purpose of global observations of the Earth and planets. The orbital altitude of polar orbits is chosen to produce a specified observation resolution and field of view. Sometimes the period must also be chosen to produce a sub-synchronous orbit thus assuring that the satellite ground track will repeat after a specified number of orbits. For Earth observation satellites, repetitive ground tracks on a weekly or monthly basis may be desirable. Since the inclination is 90°, nodal regression due to $J_2$ (equation (5-22)) is zero so to first order the orbit plane is inertially fixed. Launches into polar orbits occur from the west cost of the USA and require slightly more launch capability because the rotation of the Earth does not contribute to attaining the orbital velocity.

## 7.5 Sun-Synchronous Orbits

A **Sun-synchronous orbit** (SSO) is a nearly polar orbit where the ascending node precesses at 360 degrees per year or 0.9856 degrees per day. This type of orbit assures that the local solar time (LST) at the ascending node is nearly constant throughout the life of the mission. Orbits are identified by the time of ascending node crossing. So a "2 PM orbit" will ascend through the equator when the LST=1400 hours. Such orbits are primarily used for missions where the scientific instruments have been optimized for a particular lighting condition. Sun-synchronous orbits are typically nearly circular and always retrograde (i>90 deg). The design of a sun-synchronous orbit starts with equation (5-22) for the regression of the node

$$\frac{d\Omega}{dt} = -\frac{3}{2} n J_2 \left(\frac{R}{p}\right)^2 \cos i \qquad (7\text{-}1)$$

where for $d\Omega/dt \approx 0.9856$ degrees/day. Typical parameters for sun-synchronous orbits include an inclination between 96 and 100 degrees and as altitude of 400 -1200 kilometers.

Exercise 7-3. Make a plot of the relationship between semi-major axis and inclination for circular sun-synchronous orbits. Discuss any limiting factors for the design of such an orbit, e.g. maximum or minimum altitude. Identify, with a "+" on the plot, all combinations of a and i that will have repetitive ground tracks each day, i.e. orbits that are both sun-synchronous and sub-synchronous.

## 7.6 Low-Earth Orbit (LEO)

A low-Earth orbit is roughly defined as any orbit with an altitude less than that of a geosynchronous orbit. Almost 90 percent of all satellites in orbit are in LEO. LEO is so often utilized because of the low launch requirements that are needed to place a satellite into orbit. The highest density of satellites are at altitudes between 200 kilometers and 1000 kilometers. Altitudes below 200 kilometers are not practical because of atmospheric drag. Orbits above 1000 km often utilize a lower altitude parking orbit as in intermediate phase to attaining the final orbit. Manned spacecraft utilize LEO in the relatively radiation-free corridor above 200 km and below 600 km. LEO is used for such missions as flight tests, Earth observations, astronomical observations, space stations, scientific experiments, and possibly commercial endeavors.

## 7.7 Frozen Orbits

A **frozen orbit** is designed so that one or more of the orbital elements are "frozen" or held constant in time. In other words, the orbit is designed so that there are no secular or long period perturbations (Section 5.4.2) to specified orbital elements. One example is the Molniya orbit where the choice of inclination assures that the secular change in $\omega$ is minimized. Some missions require a nearly constant eccentricity to either extend mission life time, to provide nearly constant altitude observations, or to meet other mission requirements. The eccentricity of low altitude satellites is influenced by the odd zonal harmonics of the gravity field (5.4.4) and reducing this influence may be necessary to extend mission duration. This is particularly important for the life times of LLO. Atmospheric drag (5.5.1) and solar radiation pressure (5.4.5) also affect eccentricity but are ignored in the following. The averaged variation of parameters equation (5-24) for eccentricity changes due to $J_3$ is

$$\frac{de}{dt} = \frac{3J_3 n}{2(1-e^2)^2}\left(\frac{R}{a}\right)^3 \sin i\left(1-\frac{5}{4}\sin^2 i\right)\cos\omega \tag{7-2}$$

The eccentricity will be constant under a number of conditions: (1) argument of perigee equal to 90 or 270 degrees, or (2) inclination of 0, 63.4, 116.6, or 180 degrees. Hence if the Molniya orbits are at the critical inclination, the argument of periapsis, the eccentricity and the inclination (equation (5-26)) will have no secular or long period variations due to $J_2$ and $J_3$. The requirement of critical inclination is a rather severe restriction for general mission design and the question arises if there are other orbit conditions that lead to frozen orbits where neither eccentricity nor argument of periapsis has long term or secular variations.

Equation (5-22) includes only the $J_2$ contribution to the perturbation of $\omega$. Including the contribution of $J_3$ results in

$$\frac{d\omega}{dt} = \frac{3nJ_2R^2(4-5\sin^2 i)}{4a^2(1-e^2)^2}\left[1+\frac{J_3}{2J_2(1-e^2)}\left(\frac{R}{a}\right)\frac{\sin i\sin\omega}{e}\right] \tag{7-3}$$

The critical inclinations still provide a constant argument of perigee. Another solution occurs when the bracketed expression is zero. Further if $\omega = \pm 90$ then the eccentricity and inclination will also be constant for both $J_2$ and $J_3$ long term and secular variations. This leads to an equation for the eccentricity in terms of the semi-major axis and the inclination

$$e \approx -\frac{J_3}{2J_2}\left(\frac{R}{a}\right)\sin i \sin\left(\pm\frac{\pi}{2}\right) \tag{7-4}$$

where $\omega$ must be chosen to made the eccentricity positive. Since $J_2$ and $J_3$ for Earth, Moon, Mars, etc. are known, frozen orbit relationships can be developed between eccentricity, inclination, and semi-major axis for mission design purposes. Among these planets, the Moon has the largest ratio with $J_3/J_2$ =0.03, so that the eccentricity for these frozen orbits will be less than 0.015.

Note that equation (7-4) is not applicable near the critical inclination since equation (7-2) was derived on the assumption that $d\omega/dt \neq 0$. Near the critical inclination, the eccentricity for a frozen orbit increases significantly due to the higher order spherical harmonics that dominate near the critical inclination. The major contributors are $J_5$ and $J_7$, Reference 1 Consequently frozen orbit design near the critical inclination usually requires inclusion of higher order gravity terms and iterative numerical solutions. For a small range of inclinations, frozen orbit solutions may not exist.

## 7.8 Satellite Constellations

No single satellite can provide continuous, global coverage, so early in the utilization of space for communication and Earth observation, constellations of satellites were utilized to provide continuous coverage. NASA launched three Tracking and Data Relay Satellites (TDRS 1,3 and 4) to provide nearly global, continuous communication between other Earth satellites and almost any ground station. These satellites, which are in geostationary orbits separated by about $120^o$ in longitude, eliminated the need for NASA to maintain costly ground tracking and communication stations around the world.

Exercise 7-4. Can a TDRS communicate with a LEO as the LEO passes over the North pole? Assume that signals can not propagate through the atmosphere below 100 km. What is the minimum altitude for the LEO to maintain communications?

Since a Molniya satellite provides about 8 hours of communication opportunity for high northern latitude locations, three satellites would be required to provide continuous coverage. The three orbits could be identical except for $\Omega$ and $\tau$. The ascending nodes would be placed about $120^o$ apart and the times of periapsis would be about 8 hours apart.

The most well known constellation today is the Global Positioning Satellite system (GPS). This is a constellation of satellites that provide position and time information anywhere on the globe and was originally designed for defense purposes. The GPS system consisted of 24 satellites of which three are spares. The orbits of the spares are strategically selected so that a spare can be quickly moved into the orbit of a failed satellite. An observer's position (latitude, longitude and altitude) is determined by measuring the light time travel from the GPS satellites to the observer. If the observer's clock was synchronized with the atomic clocks on the GPS satellites, only data from three satellites would be required. Since few mobile clock maintain sufficiently accurate synchronization, a fourth satellite is necessary to synchronize the observer's clock. This requirement determines the number of satellites that make up the GPS constellation. When deployed in six equally spaced 12 hour orbits, the minimum number of satellites considered necessary to provide adequate coverage is 21. This maximizes the **constellation value**. The constellation value represents the fraction of space and time over which four satellites will be available. For the 21 satellite system, the constellation value is 0.996. Consequently at any location an observer will only have to wait a short time for 4 satellites to be in view.

The orbital elements of the satellites that make up the GPS satellite system are shown in Table 7-1. GPS is universally and routinely used to determine position to within 10 meters at any location on the Earth. In addition, it has been used for orbit determination, ionosphere mapping, spacecraft and airplane attitude determination and many other unanticipated applications.

## 7.9 Astronautics Toolbox

## 7.10 References

1. Lyddane, R.H. and C.J. Cohen, **The Low-Eccentricity Earth Satellite Orbit at the Critical Inclination,** Celestial Mechanics, Vol. 18, 1978, pg 233-236.

Table 7-1. GPS Orbital Elements

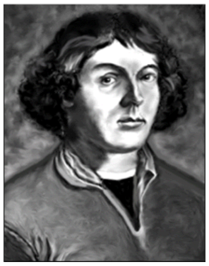
| Satellite | a (km) | e | i (deg) | Ω (deg) | ω (deg) | M (deg) |
|---|---|---|---|---|---|---|
| A1 | 26609 | 0 | 55 | 325.73 | 0 | 190.96 |
| A2 | 26609 | 0 | 55 | 325.73 | 0 | 220.48 |
| A3 | 26609 | 0 | 55 | 325.73 | 0 | 330.17 |
| A4 | 26609 | 0 | 55 | 325.73 | 0 | 83.58 |
| B1 | 26609 | 0 | 55 | 25.73 | 0 | 249.90 |
| B2 | 26609 | 0 | 55 | 25.73 | 0 | 352.12 |
| B3 | 26609 | 0 | 55 | 25.73 | 0 | 25.25 |
| B4 | 26609 | 0 | 55 | 25.73 | 0 | 124.10 |
| C1 | 26609 | 0 | 55 | 85.73 | 0 | 286.20 |
| C2 | 26609 | 0 | 55 | 85.73 | 0 | 48.94 |

Table 7-1. GPS Orbital Elements

| Satellite | a (km) | e | i (deg) | $\Omega$ (deg) | $\omega$ (deg) | M (deg) |
|---|---|---|---|---|---|---|
| C3 | 26609 | 0 | 55 | 85.73 | 0 | 155.08 |
| C4 | 26609 | 0 | 55 | 85.73 | 0 | 183.71 |
| D1 | 26609 | 0 | 55 | 145.73 | 0 | 312.30 |
| D2 | 26609 | 0 | 55 | 145.73 | 0 | 340.93 |
| D3 | 26609 | 0 | 55 | 145.73 | 0 | 87.06 |
| D4 | 26609 | 0 | 55 | 145.73 | 0 | 209.81 |
| E1 | 26609 | 0 | 55 | 205.73 | 0 | 11.90 |
| E2 | 26609 | 0 | 55 | 205.73 | 0 | 110.76 |
| E3 | 26609 | 0 | 55 | 205.73 | 0 | 143.88 |
| E4 | 26609 | 0 | 55 | 205.73 | 0 | 246.11 |
| F1 | 26609 | 0 | 55 | 265.73 | 0 | 52.42 |
| F2 | 26609 | 0 | 55 | 265.73 | 0 | 165.83 |
| F3 | 26609 | 0 | 55 | 265.73 | 0 | 275.52 |
| F4 | 26609 | 0 | 55 | 265.73 | 0 | 305.04 |

# Appendix A - Biographical Bullets

## Copernicus



Nicolaus Koppernik (1473-1543).

•Doctor of Arts and Medicine, professor of mathematics, and canon at Frauenburg.

•Compiled tables of the planetary motions that remained state of the art until Tycho Brahe.

•In 1507 became convinced of the heliocentric nature of the solar system.

•Aware that the planets speeds varied throughout the orbit.

•Knew that precession is a conical motion of the earth's axis.

- Suggested universal gravity: "that gravity is not an influence of the whole earth, but is a property of its substance, which, it is thinkable, may extend to the sun, moon, and other stars."[1]
- First treatise on the heliocentric theory written about 1530. Pope Clement VII approved the work and asked for a complete presentation, which was finished in 1540. Printing in 1543 was paid for by a cardinal and presented to Copernicus on his death-bed. It was dedicated to Pope Paul III. He did not fear religious criticism: "If there be some babblers," he wrote, "who, though ignorant of mathematics, take upon them to judge of these things, and dare to blame and cavil at my work, because of some passage of Scripture which they have wrested to their own purpose, I regard them not, and will not scruple to hold their judgment in contempt." During his lifetime his work received the approval of the Church.[1]
- In 1616, Copernician theory was declared "false, and altogether opposed to Holy Scripture," and placed on the *Syllabus Errorum*.
- At the end of the sixteenth century the theory of Copernicus was warmly, if not hotly, upheld by Giordano Bruno, who welcomed its aid in his attack on Aristotle. Bruno was imprisoned, excommunicated, and burned at the stake in 1600, and scientific men of succeeding generations cannot have been unmindful of his fate.[1]

## Euler

Leonhard Euler (1707-1783)

•Founded analytical mechanics in *Mechanica sive Motus Scientia* (1736).

•Initiated calculus of variations with development of the “Euler-Lagrange” equation in 1744.

•First to provided detailed analytic treatment of the two body problem in *Theoria motuum planetarum et cometarum* (1744) vs geometric approach.

•Provided analytic solution to orbit determination problem (1744).

•Applied the variation of parameters method to the study of the mutual perturbations of Jupiter and Saturn and received the French Academy of Sciences award in 1748 and 1752. Also used to calculate the effects of drag on projectiles.

- Knew of the ten integrals of the n-body problem.
- Proposed “center of inertia” to distinguish from “center of gravity.”
- Studied bending of beams, motion of fluids, columns in compression, blood flow, refraction and dispersion of light, etc. etc.
- Received 300 pounds British from the Longitude Act for his analytic theory of lunar motion, 1764.[2]

- Published two lunar theories (1753 and 1772). The first based on variation of parameters and the second referenced to the mean rotation of the moon.
- Wrote 400 pages of his books while totally blind [3].

## Galileo

Galileo Galilei (1564-1642)
•Born Florence and died the year of Newton's birth.
•At 17 he was a student of medicine.
•At 26, he was a professor of mathematics, having been seduced by the mathematics of Euclid and Archimedes.
•Experimented with falling bodies including the famous Leaning Tower experiment and inclined planes reaching the conclusions:

1. The final velocity acquired is independent of the angle of slope, but depends on the vertical height through which the body falls.
2. The height through which the body falls is proportional to the square of the time.
3. All bodies fall at the same rate.

- Received from Kepler a copy of his *Mysterium Cosmographicum* and wrote "I have been for many years an adherent of the Copernican system. I have collected many arguments for the purpose of refuting [the commonly accepted hypothesis], but I do not venture to bring them to the light of publicity, for fear of sharing the fate of our master Copernicus, who has become the object of ridicule and scorn. I should certainly venture to publish my speculations if there were more people like you. But this not being the case, I refrain from such an undertaking."
- Realized that the new star of 1604 ended the Aristotelian view of the heavens and he began publicly supporting the Copernican theory[1]
- Discovered the telescope and used the 3x refractor for observing planets and stars. Eventually built a 30x telescope.
- Discovered now familiar lunar surface features and estimated the heights of lunar mountains.
- Recognized the earth is brighter than the moon because of the clouds.
- Observed the four brightest moons of Jupiter, and demonstrated that they were in orbit around the planet.
- Proposed and diligently pursued using the eclipses of the Jovian moons to solve the longitude problem and win the life pension prize offered by King Philip III [2]
- Observed in 1610 that Saturn was a triple planet.
- Observed in 1612 that Saturn was a single planet and commented: "Looking at Saturn within these last few days, I found it solitary and without its accustomed stars, and in short perfectly round and defined like Jupiter. Now what can be said of so strange a metamorphosis? Are perhaps the two smaller stars consumed like spots on the sun? Have they suddenly vanished and fled? Or has Saturn devoured his own children? Or was the appearance indeed fraud and delusion? . . . Now perhaps the time is come to revive the withering hopes of those who . . . have fathomed all the fallacies of the new observations and recognized their impossibility . . . The shortness of time, the unexampled occurrence, the weakness of my intellect, the terror of being mistaken, have greatly confounded me."[1]

- Discovered the phases of Venus, and announced them in the form of an anagram: "the mother of the Loves (Venus) imitates the phases of Cynthia (the moon)." Thus one of the objections to the heliocentric hypothesis was removed. [1]
- Slight phase changes were even correctly observed for Mars.
- Wrote for the church an apparently dispassionate dialogue that compared the Ptolemaic and Copernican systems; however, the astronomical arguments for the latter that it contained were cogent and unanswerable, and the book was suppressed.[1]
- Put on trial for his books, which he had already twice revised. Made public submission: "I do not hold, and have not held this opinion of Copernicus since the command was intimated to me that I must abandon it . . . I swear that in future I will never say or assert, verbally or in writing, anything that might furnish occasion for a similar suspicion against me." [1]

## Gauss

Carl Friedrich Gauss (1777-1855)
•Developed the least squares method.
•Perhaps the first to recognize the concept of radius of convergence for power series.[3]
•Developed orbit determination method for Earth based observations, 1801[4].
•Introduced hypergeometric functions in 1812, *Disquisitiones Generales Circa Seriem Infinitam*.
•1999 study of his brain indicated nothing unusual, unlike Einstein's that has unusually large inferior parietal lobes.

## Jacobi

Carl Jacobi (1804-1851)
•Proved that if all but the last two integrals of the n-body problem were known, then the last two could be found (1842).
•Showed that the n-body problem was equivalent to solving a partial differential equation of one-half the order of the original ordinary differential equations, setting the stage for the Hamilton-Jacobi theory.
•Reduced the general problem of three bodies to seventh order (1843), no further reduction to date.

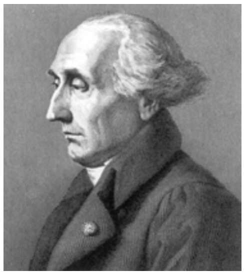
## Kepler

Johannes Kepler (1571-1630)
•Twenty-five years younger than Tycho, and seven years younger than Galileo.
•In his twenties he was interested in the planetary motions, and published a theory based on the forms of the regular solids.
•Fuller knowledge of the facts destroyed this formal picture after he began to work with Tycho in 1599.[1]
•Worked primarily with Tycho's observations of Mars, and he devoted enormous labor to improving the Copernican picture of circular motion in an orbit not quite centered on the sun.

- Announced the first two of his three laws of planetary motion in 1609, and the third in 1618.

- Conjectured in a 1611 pamphlet, “The Six Cornered Snowflake,” the optimal spacing for spherical bodies is the face centered cubic, finally proved in 1998 to be true.
- Derived much of his income from computing horoscopes and astronomical almanacs.
- Successfully defended his mother during her trial for witchcraft (1615-21).[1]

## Lagrange



Joseph-Louis Lagrange (1736-1813)
•First memoir on perturbations of Jupiter and Saturn, included further development of the variation of parameters method, 1766 [4]. Inclination, node and longitude of perihelion were correct, but incorrectly assumed that semi-major axis and time of perihelion were constant. Mean longitudes included secular terms proportional to time and time squared (see Laplace)
•Presented particular solutions to the problem of three bodies, 1772
•Showed in 1776 that there are no first order secular variations in the semi-major axes of the planets for all orders of eccentricity.[4]

- In a prize memoir, variation of parameters method developed as we know it today(1782) and applied to the perturbation of elliptical comets.
- Developed power series expansion in time and introduced recursive variables for coefficients leading to F and G functions.

## Laplace

Library of Congre

Pierre-Simon de Laplace (1749-1827)
•Devoted most of his life to the study of celestial mechanics and his *Mecanique Celeste* (1799-1805) contained all that was known about the subject at that time.
•First memoir to the French Academy of Sciences (1773) proving that up to second powers of the eccentricity, the semi-major axes of the planets had no secular variations.[4]
•Conceived of the invariant plane in 1784.
•Explained the secular acceleration of the moon’s mean motion (1787).

- Showed that outside a gravitational body, $\nabla^2 V = 0$. Green named V the potential function in 1828.
- Demonstrated that secular terms in mean longitude were actually long period variations [4].
- Developed the planetary disturbing function to third order in eccentricity and inclination.
- Developed fundamentally new orbit determination method (1780).

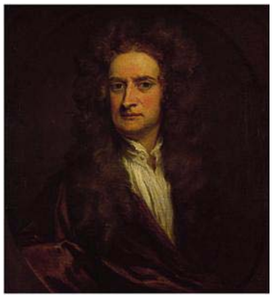
## Napier

Napier, John (1550-1617)
•Invented "artificial numbers", we call them logrithms, in 1614.
•Invented Napier’s bones, a set of marked wooden pegs for doing arithmetic, later became the slide rule.
•Tycho Brahe waited in vain for Napier to complete his log tables and both died before they were completed.
•Henry Briggs first completed the tables and for a long time they were named for him.

## Newton

Nature and nature's law lay hid in night
God said, Let Newton be, and all was light

POPE, from the inscription on Newton's monument



Isaac Newton (1642-1727). In the words of Leibnitz: "Taking mathematics from the beginning of the world to the time when Newton lived, what he had done was much the better half." [4]

- His life appeared to be uneventful and that of a retiring scholar.
- Most of his important scientific work was done in his earlier years.
- In his old age he wrote at great length on theological problems, much of which remains unpublished even though Newton thought it his most important production [1].
- Reluctant to go to the trouble of publishing his results, and the *Principia* would probably never have appeared if Halley had not paid for it and seen it through the press [1].
- The *Principia*, containing the derivations of the laws of planetary motion, is a formulation of celestial mechanics that has not been essentially improved up to the present.
- *Principia* is difficult for the modern reader because all the proofs are geometrical. For a modern geometrical treatment see [5].
- Proves that for central forces, the law of areas (Kepler's second law) must be obeyed, whatever the nature of the force, and that if the law of areas is obeyed, the force must be central.
- Examined the law of force that will produce motion in an elliptical orbit round the center of the ellipse and showed that in this case the force will vary directly as the distance from the center and the period is independent of the size of the ellipse (the 3-d harmonic oscillator)[4].
- Investigated motion under a central force in an ellipse about the focus, and shows that it implies an inverse square law. He goes on to prove that motion in confocal ellipses under an inverse square law of attraction will result in periodic times that are as the 3/2 powers of the major axes (Kepler's third law).
- Extends the proof to the hyperbola and the parabola.
- To apply the gravity law to bodies of finite size, proved that a sphere attracts as though its mass we're concentrated at its center.
- Considers how to determine orbits. A conic is fixed by five points through which it passes; however, the determination of the motion of a planet in a conic requires only three, because the law of motion supplies the equivalent of the other two.[4]
- Discusses the motion of the moons of Jupiter and Saturn, and that of the planets.
- Gives the theory of the figure of the earth and of the tides.
- Makes a dynamical study of precession.
- Invent the "Method of Fluxions," essentially modern calculus.
- He did not publish his results at once. It was not until 1684, when Wren offered a small prize for the first who should prove that a body under the inverse square law would describe an ellipse, that Halley spoke to Newton of the problem, and found that he had solved it sixteen years earlier! Halley persuaded Newton to publish, and the *Principia* appeared in 1687.[4]

## Poincare

Henri Poincare (1854-1912)
•Demonstrated that if two mass are small in the problem of three bodies, there are an infinity of periodic solutions.
•Fundamentally new approach to celestial mechanics published in *Les Methods Nouvelles de la Mechanique Celeste*.
•Proved that rotating fluids can have an infinity of equilibrium shapes.
•Discovered the dynamical concept we now call “chaos.”

## Ptolemy

Ptolemy (~100-170 AD)
•Wrote *Almages*t describing his works which has remained intact to modern times.
•Described the evection in the Moon’s motion.
•Discovered refraction of light.
•Used eccentricities and epicycles to explain the apparent motion of the planets.
•Last significant astronomical (except for the Arabs) research until the end of the dark ages.

## Tisserand

Francois-Felix Tisserand. (1845-1896)
• *Traite’ de Mechanique Celeste*, 1891-1896, outstanding four volume work provides comprehensive coverage of gravity potentials for irregular bodies, stability of the solar system, development of perturbation functions, etc.
• Doctorial thesis extended Delaunays research on the three body problem for the lunar theory.
•
•

## Tycho Brahe

Tycho Brahe, (1546-1601)
•First modern astronomer. Began construction of quadrants and sextants at early age after seeing an eclipse of the sun.
•New star of 1572 stimulated him to build an observatory on the island of Hven, off the Swedish coast, in 1576.
•Undertook study of planetary motions, and constructed the most accurate instruments that were possible at the time. He took great precautions in making observations and was thoroughly modern in his attempts to avoid and evaluate errors. His magnificent series of planetary observations made possible Kepler's study of the laws of planetary motion, the basis of our modern picture of the solar system.[4]

- Did not accept the heliocentric ideas of Copernicus. Believed that the earth was the stationary center of the solar system; its motion, he argued, is not felt and is difficult to picture. Moreover, he saw that the stars should show annual parallax if the earth has orbital motion; he did not realize that his failure to detect the parallax was a result of the great distance of the stars and the consequent extreme smallness of the effect. He pointed out (as Copernicus had done) that Mercury and Venus should show changing phases if the sun is at the center of the system; we now know that they display such phases, but Tycho could not detect them. He argued that if the earth is in motion, a stone should not fall vertically; we know now that it does not. His picture of the solar system showed the sun, moon, and superior planets carried around the earth, but with Mercury and Venus going in orbits around the sun (Heracleides of Pontus had held the same view two millenniums before) [4].
- Published works include astrological predictions (which were reputable in the sixteenth century) and accounted for much of his revenue.
- Lost his nose and his honor in a duel after astrologically predicting the death of an already dead sultan.
- Late in life he was assisted by Johann Kepler.

## REFERENCES

1. Payne-Gaposchkin, Cecilia, *Introduction to Astronomy,* Prentice-Hall, Inc. New York, 1954.
2. Sobel, Dava, *Longitude*, Penguin Books, 1995
3. Battin, Richard, *An Introduction to the Mathematics and Methods of Astrodynamics*, AIAA Education Series, 1987.
4. Moulton, Forest, *An Introduction to Celestial Mechanics*, MacMillan Company, 1914. Also available from Dover Publications, Inc.
5. Goodstein, D. L. & J. R. Goodstein, *Feynmans's Lost Lecture - The Motion of Planets Around the Sun*, Norton & Company, 1996.

# Appendix B - Vector and Matrix Analysis

This appendix provides some useful formulae from vector and matrix analysis without proof or expanded explanation. Since the applications for this book are generally to position and velocity vectors, the equations are valid for real, 3 vectors and 3 by 3 matrices. Some of the equations may also be valid for more general linear vector spaces. In the equations below regular text variables represent scalars, bold lower case represent 3 vectors and bold upper case represent 3 by 3 matrices.

## B-1 Algebra

**Unit vectors**: $\mathbf{e}_g$ is a unit vector in direction g. Typical replacements for g are x, y, z, r, $\phi$, $\theta$,...

**Vectors**: $\mathbf{a} = (a_1, a_2, a_3) = (a_x, a_y, a_z) = a_x\mathbf{e}_x + a_y\mathbf{e}_y + a_z\mathbf{e}_z$

**Matrices**: $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$

**Vector norm or length**: $a = |\mathbf{a}| = \sqrt{{a_x}^2 + {a_y}^2 + {a_z}^2}$

**Addition**: Is component wise, e.g. $\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$. Likewise for matrices.

**Commutative law of addition**: $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ and $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

**Associative law of addition**: $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$ and $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$

**Vector dot or inner product**: $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} = |\mathbf{a}||\mathbf{b}|\cos(\mathbf{a},\mathbf{b}) = a_x b_x + a_y b_y + a_z b_z$ Can be thought of as the length of **a** times the length of **b** projected on to **a** or conversely. Commonly used to determine the cosine of the angle between two vectors which is always between 0 and $\pi$.

**Orthogonal vectors**: **a** and **b** are orthogonal iff $\mathbf{a} \cdot \mathbf{b} = 0$

**Vector cross or outer product**: $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} = (a_y b_z - a_z b_y,\ a_z b_x - a_x b_z,\ a_x b_y - a_y b_x)$ Commonly used to determine the sine of the angle between two vectors since $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}|\sin(\mathbf{a},\mathbf{b})$. Be aware of quadrant limitations.

**Distributative laws**: $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$ and $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ and $\mathbf{A(B+C) = AB + AC}$

**Triple scalar product**: $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \mathbf{c} \cdot \mathbf{a} \times \mathbf{b} = \mathbf{b} \cdot \mathbf{c} \times \mathbf{a}$ is preserved under cyclic permutation or operator interchange $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c}$, but changes sign if two terms are interchanged $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = -(\mathbf{b} \cdot \mathbf{a} \times \mathbf{c}) = -(\mathbf{a} \cdot \mathbf{c} \times \mathbf{b})$. The triple scalar product is equal to the volume of the parallelepiped form by the three vectors as sides.

**Triple vector product**: $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$

**Matrix product**: **C=AB** where the ij term of **C** is given by $C_{ij} = \sum_{k=1}^{3} A_{ik}B_{kj}$

**Matrix-Vector product**: **c**=**Ab** if given by $c_i = \sum_{j=1}^{3} A_{ij}b_j$ maps the vector **b** into vector **c.**

**Matrix transpose:** $\mathbf{A}^T$ or **A'** If $A_{ij}$ is the ij-th term of **A** then $A_{ji}$ is the ij-th term of $\mathbf{A}^T$

**Matrix identity:** The 3 by 3 matrix multiplicative identity is $\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

**Matrix inverse:** If the 3 by 3 matrix **A** has an inverse $\mathbf{A}^{-1}$ then $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_3$.

**Orthonormal matrix:** **A** is orthonormal iff $\mathbf{A}^T = \mathbf{A}^{-1}$**.**

**Rotation matrix**: An orthonormal (rows and columns are orthogonal and have unit length) matrix that describes the orientation between two orthogonal coordinate systems. For example a rotation through angle $\alpha$ about the z-axis produces the rotation matrix $\begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$

## B-2 Calculus

**Time derivatives**:
Time derivative of a vector is the time derivative of the components

$$\dot{\mathbf{a}} = \frac{d\mathbf{a}(t)}{dt} = \left(\frac{da_1}{dt}, \frac{da_2}{dt}, \frac{da_3}{dt}\right)$$

$$\frac{d(\mathbf{a} \cdot \mathbf{b})}{dt} = \frac{d\mathbf{a}}{dt} \cdot \mathbf{b} + \mathbf{a} \cdot \frac{d\mathbf{b}}{dt} \qquad \frac{d(\mathbf{a} \times \mathbf{b})}{dt} = \frac{d\mathbf{a}}{dt} \times \mathbf{b} + \mathbf{a} \times \frac{d\mathbf{b}}{dt}$$

For any vector $\mathbf{a} \cdot \frac{d\mathbf{a}}{dt} = |\mathbf{a}|\frac{d|\mathbf{a}|}{dt}$. So if $\mathbf{a}$ is a unit vector, the dot product vanishes, i.e. the time derivative of a unit vector is orthogonal to the unit vector.

**Spatial derivatives**:

**Gradient** of a scalar function: $\nabla f(x, y, z) = \frac{\partial f}{\partial x}\mathbf{e}_x + \frac{\partial f}{\partial y}\mathbf{e}_y + \frac{\partial f}{\partial z}\mathbf{e}_z$ is a vector normal to the surface f=0 and pointing in the direction of increasing f

**Divergence** of a vector field: $\mathrm{div}\ \mathbf{a} = \nabla \cdot \mathbf{a}(x, y, z) = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}$

**Curl** of a vector field

$$\mathrm{curl}\ \mathbf{a}(x,y,z) = \nabla \times \mathbf{a} = \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right)\mathbf{e_x} + \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right)\mathbf{e_y} + \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)\mathbf{e_z}$$

$\nabla \times \nabla f(x, y, z) = \mathbf{0}$ which is the basic for the theorem that a force is derivable from a potential function iff the curl of the force field vanishes everywhere.

**Laplacian:** $\nabla \cdot \nabla f(x, y,z) = \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$

# Appendix C - Acronyms and Notation

## Notation

**v** - - - - vectors are bold characters
$\mathbf{e}_b$ - - - - unit vector in direction b, typical replacements for b are x, y, z, r, $\phi$, $\theta$, $\xi$, $\eta$, $\zeta$

## Acronyms

DOY - - - day of year
EOM - - equation of motion
ET - - - - ephemeris time, (1.3.3)
JD - - - - Julian day number, (1.3.4)
MDJ - - - modified Julian day number= JD-2400000.5, (1.3.4)
TAI - - - atomic time, (1.3.1)
TDB - - - barycentric dynamic time, (1.3.2)
TDT - - - terrestrial dynamic time, (1.3.2)
UT - - - - universal time, (1.3.6)
wolog - - without loss of generality
wrt - - - - with respect to

# Index