Orbital Mechanics Course Notes

David J. Westpfahl
Professor of Astrophysics,
New Mexico Institute of Mining and Technology

March 31, 2011

These are notes for a course in orbital mechanics catalogued as Aerospace Engineering 313 at New Mexico Tech and Aerospace Engineering 362 at New Mexico State University. This course uses the text "Fundamentals of Astrodynamics" by R.R. Bate, D. D. Muller, and J. E. White, published by Dover Publications, New York, copyright 1971. The notes do not follow the book exclusively. Additional material is included when I believe that it is needed for clarity, understanding, historical perspective, or personal whim.

We will cover the material recommended by the authors for a one-semester course: all of Chapter 1, sections 2.1 to 2.7 and 2.13 to 2.15 of Chapter 2, all of Chapter 3, sections 4.1 to 4.5 of Chapter 4, and as much of Chapters 6, 7, and 8 as time allows.

Purpose

The purpose of this course is to provide an introduction to orbital mechanics. Students who complete the course successfully will be prepared to participate in basic space mission planning. By basic mission planning I mean the planning done with closed-form calculations and a calculator. Students will have to master additional material on numerical orbit calculation before they will be able to participate in detailed mission planning.

There is a lot of unfamiliar material to be mastered in this course. This is one field of human endeavor where engineering meets astronomy and celestial mechanics, two fields not usually included in an engineering curriculum. Much of the material that is familiar to students of those disciplines will be unfamiliar to engineers. Students are probably already familiar with Newton's Laws and Newtonian gravity. These will be used to develop the particular applications needed to describe orbits and orbital maneuvers.

Space missions are expensive and risky, especially if people or living animals are sent into space. Thus, it is important to check and recheck calculations and assumptions. Computer programs are subject to the imperfections of the humans who write them. This, it becomes necessary to develop physical insight into orbit calculations to have a sense of when a programming bug is leading to inaccurate answers. We will spend time developing physical intuition and understanding what it is.

Notation

Well I remember being a student and being frustrated by notation used by printers of textbooks that was impossible to write by hand at note-taking speed, and not used by the professor, anyway. Thus, I have tried to use notation that is consistent with the text, but also within my abilities to write on the board and within the abilities of students to write in their notes. Some examples follow.

A scalar is written as an ordinary math symbol, as in a.

A vector is written with an arrow above, as in $\vec{r}$.

A unit vector is written with a hat, as in $\hat{r}$. If the unit vector is a basis vector of a coordinate set it's symbol is usually capitalized, as in $\hat{I}$.

A matrix is written in a boldfaced capital letter and covered by a tilde, as in $\tilde{\mathbf{D}}$. This is a compromise. The tilde is easy enough to write in notes or on the board, but boldface is not. The boldface is used to make the matrix instantly recognizable in the notes, at the cost of inconsistency.

The inverse of the same matrix is written as $\tilde{\mathbf{D}}^{\mathbf{-1}}$.

The transpose of a matrix is written in boldface with a tilde and a trailing superscript capital T, as in $\tilde{\mathbf{D}}^{\mathbf{T}}$.

A triangle with vertices A, B, and C is named ABC. Its line segments are named AB, BC, and CA. Order does not matter, so AB and BA describe the same line segment. The angle between segments AB and BC is labeled $\widehat{ABC}$.

These notes were made using the LaTeX math symbols of AMS TeX. Anyone who has posted hundreds of pages of LaTeX notes has probably discovered hundreds of typos and left undiscovered scores of others. I am no exception. If you discover typos please report them to me by email at dwestpfa@nmt.edu.

D. J. W.
Albuquerque, NM
January, 2011

Contents

1 Two-Body Orbital Mechanics **1**

1.1 Kepler's Laws . . . 1

1.2 Newton's Laws . . . 2

1.2.1 An Example . . . 4

1.2.2 History . . . 8

1.3 The Equation of Motion for Two Orbiting Bodies . . . 9

1.3.1 Choosing a Goal . . . 9

1.3.2 Making Things Complicated . . . 10

1.3.3 Making Things Simple Again . . . 10

1.4 Partial Solutions of the Equation of Motion . . . 14

1.4.1 Constants of the Motion . . . 14

1.4.2 The Trajectory Equation . . . 17

1.5 Conic Sections . . . 18

1.5.1 Polar Equations of Conics . . . 19

1.6 Properties of Conic-Section Orbits . . . 24

1.6.1 Relating the Constants of the Motion to the Geometry of the Orbit . . . 26

1.6.2 Some Important Properties of Individual Conic Orbits 27

1.6.3 Relationships Among the Conics . . . 30

1.6.4 What is $\vec{B}$? . . . 32

1.6.5 The Eccentricity Vector . . . 34

1.7 Canonical Units . . . 34

2 Orbit Determination from Observations **39**

2.1 Coordinate Systems . . . 39

2.1.1 Heliocentric-Ecliptic Coordinates . . . 40

2.1.2 Geocentric-Equatorial Coordinates . . . 41

2.1.3 Right Ascension-Declination Coordinates . . . 41

2.1.4 Perifocal Coordinates 42
2.2 Classical Orbital Elements 42
2.3 Determining the Orbital Elements from $\vec{r}$ and $\vec{v}$ 44
2.3.1 Three Fundamental Vectors 44
2.3.2 Solving for the Orbital Elements 45
2.4 Determining $\vec{r}$ and $\vec{v}$ from the Orbital Elements 49
2.5 Coordinate Transformations 51
2.5.1 Transformations Change the Basis Vectors 51
2.5.2 Simple Transformations 52
2.5.3 More Challenging Transformations 55
2.6 Mechanics on the Rotating Earth 63
2.6.1 An Introduction to Time 63
2.6.2 Position, Velocity, and Acceleration 67
2.7 The Ellipsoidal Earth 72
2.7.1 The Measurement of Latitude 73
2.7.2 Station Coordinates 73
2.8 The Ground Track of a Satellite 85
2.8.1 Launch Site, Launch Azimuth, and Orbital Inclination 85

3 Real Orbits and Orbital Maneuvers 89
3.1 Some Types of Orbits 89
3.1.1 Classification of Earth Orbit by Altitude 89
3.1.2 Classification of Orbits by Inclination 91
3.1.3 Sun-Synchronous Orbits 92
3.1.4 Special Orbits 92
3.2 The Earth's Equatorial Bulge 93
3.3 In-Plane Orbit Changes 94
3.3.1 Launching a Satellite and Adjusting its Orbit 94
3.3.2 Hohmann Transfer 96
3.3.3 General Coplanar Transfer 100
3.4 Bi-elliptic Transfer 106
3.5 Out-Of-Plane Orbit Changes 111

4 $\vec{r}$ and $\vec{v}$ as Functions of Time 113
4.1 What we Have Done 113
4.2 Elliptical Time of Flight as a Function of E 114
4.2.1 Two Approaches: Geometric and Analytical 114
4.2.2 Why is This Integral so Difficult? 116

4.2.3 Kepler's Geometric Method - Developing Kepler's Equation . . . 117
4.2.4 Time of Flight Between Arbitrary Points . . . 121
4.2.5 Analytical Method . . . 123
4.3 Parabolic Time of Flight as a Function of D . . . 125
4.4 Hyperbolic Time of Flight as a Function of F . . . 127
4.4.1 Area as a Measure of Angle . . . 127
4.4.2 Hyperbolic Time of Flight . . . 131

5 Orbit Determination from Two Positions and Time 137

6 Ballistic Missile Trajectories 139
6.1 History . . . 139
6.2 Purpose . . . 140
6.3 Basics of Ballistic Missiles . . . 140
6.3.1 The Non-dimensional Parameter, Q . . . 141
6.3.2 The Free-Flight Range Equation . . . 141
6.3.3 The Flight-Path Angle equation . . . 142

7 Lunar Trajectories 145
7.1 Sphere of Influence . . . 145
7.2 The Patched Conic Approximation . . . 146

8 Interplanetary Trajectories 147

Chapter 1

Two-Body Orbital Mechanics

A story has to start somewhere. Our story starts with Kepler's Laws.

1.1 Kepler's Laws

Following our text, Fundamentals of Astrodynamics by Bate, Mueller, and White, we start with Kepler's Laws of Planetary Motion, which are generalizations derived from the planetary position data of Tycho Brahe. According to our text, Kepler published the first two laws in 1609, and the third in 1619.

Kepler's Laws

Kepler's First Law - The orbit of each planet is an ellipse with the Sun at one focus.

Kepler's Second Law - The line joining any planet to the Sun sweeps out equal areas in equal times.

Kepler's Third Law - The squares of the periods of any two planets are in the same proportion as the cubes of their mean distances from the Sun.

These laws explain what the orbits are. Their shapes are ellipses, and the local, instantaneous speeds within the ellipses change so that the second and third laws are true. These laws do not explain why the shapes and speeds of the orbits behave in this way. Our first goal will be to apply the work of Newton to understand why.

1.2 Newton's Laws

Newton did his fundamental work in the 1660's, but did not publish until the Principia appeared in 1687. Without the help of Edmund Halley the Principia might never have seen the light of day. The Principia contains Newton's three Laws of Motion and the Law of Universal Gravitation (more on that later). You will find variations in the statements of the Laws of Motion because the Principia is in Latin, and there is more than one English translation.

Newton's Laws of Motion

Newton's First Law - A body at rest stays at rest, and a body in motion stays in uniform, straight-line motion, unless acted upon by a net force.

Newton's Second Law - The time rate of change of momentum is proportional to the impressed force and is in the same direction as that force, or $\frac{d\vec{p}}{dt} = \vec{F}_{tot}$.

Newton's Third Law - For every action there is an equal and opposite reaction, *or*, forces always come in pairs. If body A exerts a force on body B, then body B exerts a force on body A that is equal in magnitude and opposite in direction.

The First Law corrects a fallacy in Aristotle's physics, that the natural state of motion is rest. Newton says that the natural state of motion is straight-line, uniform motion, that is, motion with constant linear momentum. Rest is a special case in which the momentum is zero. It is interesting that most students naturally believe in Aristotle's physics, even if they have never heard of it. They work hard in Physics 121 to overcome this natural belief.

The Second Law is effectively a definition of force from momentum, or from mass and acceleration. It explicitly involves the time rate of change, so it assumes that the reader is familiar with calculus. Let us remind ourselves of the definition of momentum,

$$\vec{p} = m\vec{v},$$

so that

$$\frac{d\vec{p}}{dt} = \frac{d(m\vec{v})}{dt} = m\frac{d\vec{v}}{dt} + \vec{v}\frac{dm}{dt} = \vec{F}_{tot},$$

where $\vec{F}_{tot}$ is the total force.

Force and momentum are vectors. This is not stated explicitly in the Second Law, I have added that part. Newton does say that the direction of the momentum change and the impressed force are the same, implying a vector relationship. The term *impressed force* is a bit vague, and some impressed forces can be counteracted, at least partially, by friction, which may be passive. Some statements of the Second Law use the phrase *net force* or *unbalanced force*, which may impart more clarity. The force that causes a change in momentum is one that is *not* counteracted or nullified by other forces.

The fact that forces are vectors is under appreciated, and its proof is too easily ignored. The statement that the time rate of change of momentum equals the total force implies a vector sum of forces.

The fact that $\vec{F}$, $\frac{d\vec{p}}{dt}$, and $m\vec{a}$ are all vectors means that they can, and, in application, must, be resolved into components. In every application of Newton's Laws we will require ourselves to choose and explicitly state a coordinate system, unless there is a compelling reason not to. We will make frequent use of Cartesian coordinates and spherical polar coordinates. One of our challenges will be to develop the definition of what a coordinate transformation is and what it does.

The Third Law turns out to be the one that challenges students the most, even more than the First Law. This is because, while forces do come in pairs, the two forces act on different bodies. This must be fully understood before the Second Law is applied. We use free body diagrams as a bookkeeping technique to assure that the bodies on which the forces act are properly assigned.

We need to work at least one example, but first a reminder. Bodies may exert forces on themselves, but these forces do not cause a change in momentum. Consider the classic example of pulling one's self up by one's bootstraps. If you pull up on your bootstraps, your bootstraps must pull downward on you. By the Third Law these forces must be equal and opposite. So far, so good. Assuming that your boots do not rip apart and that your boots are are attached to your feet, the forces cancel and no change in momentum results.

When do forces result in change in momentum? Forces cause a change in momentum when the total force or net force is nonzero, $\frac{d\vec{p}}{dt} \neq 0$ when $\vec{F}_{tot} \neq 0$.

Problem 1 The Third Law

If you cannot exert a net force on yourself how can you stand up from your chair at the end of class?

1.2.1 An Example

As an example, consider a very simple railroad train, consisting of one engine, one freight car, and one caboose. Let the train operate on a straight, level track. We choose a two-dimensional coordinate system with x along the track and y upward, perpendicular to the track. By convention we must choose a right-handed coordinate system, so let x be to the right. Let us assume that the train cars individually are of constant mass, so we know from the Second Law that

$$\frac{d\vec{p}}{dt} = \vec{F} = m\vec{a}. \tag{1.1}$$

This is a vector equation, so a similar equation must apply in each coordinate direction,

$$F_x = ma_x, \tag{1.2}$$

and

$$F_y = ma_y. \tag{1.3}$$

There is no motion in the z direction, so we may choose to ignore the application of the Second Law there. If we do apply the Second law to the z direction the result is trivial because $F_z = 0$ and $a_z = 0$. Even with this simplification we will see that a thorough analysis of the motion of the train becomes complex very rapidly.

By convention, we draw a free-body diagram for each of the three railroad cars. The argument above shows that we need to be concerned with the x and y directions only. Consider the engine first. At least four forces act on the engine. Three are obvious - the engine's weight due to gravity, the upward force of the track on the engine (in the y direction), and the forward or backward force of friction of the track on the engine (in either the plus or minus x direction). The engine and track are in contact where the wheels touch the track, so that is where the friction and the upward force act. The friction may be forward or backward according to what the engineer is doing with the drive motor and the brakes. There must be some friction or the train could not move. Judicious application of the power of the drive motor must move the train forward, similar use of the brakes must slow it down.

Let's assume that the engineer is controlling the power so that the train is moving forward.

The fourth force is not so obvious. The engine exerts a force on the freight car, thus pulling the freight car forward, so the freight car must exert a backward force on the engine according to the Third Law. This is the fourth force. Thus, the free-body diagram of the engine shows four forces: the weight downward, the track normal force upward, a frictional force from the track that is forward, and a backward force exerted by the freight car on the engine.

Next, consider the freight car. It, too, has four forces in its free-body diagram. There is still a weight downward and a track force upward. The engine pulls the freight car forward. The freight car is attached to the caboose, and the freight car pulls the caboose forward, so by the Third Law the caboose must exert a backward force on the freight car. These are the four forces. We could include a fifth force, the friction of the track on the car, but we choose to ignore it.

Finally, consider the caboose. It has only three forces in its free-body diagram, because it is not attached to anything behind it. There is still a weight, still a track force upward, and the force exerted by the freight car on the caboose. Again, we choose to ignore friction.

We have all watched or ridden in trains that operate on a track that is essentially straight and level. Such a train never sinks into the Earth or flies upward from the track, so we infer that a_y is always zero, or very nearly so. By application of the Second Law, this means that the sum of the forces in the y direction must also be zero, so the weight of the engine and the upward force on the engine exerted by the track must be balanced - they must be equal and opposite, so their sum is zero. This conclusion applies to all of the cars in the train. It means that all of the interesting things that happen are in the x direction.

Apply the Second Law to the engine using its free-body diagram as a guide. In the positive x direction we have the force of friction on the engine, f_e. In the negative x direction we have the force of the freight car on the engine, F_{fce}. The Second Law gives

$$f_e - F_{fce} = m_e a_e, \tag{1.4}$$

where m_e is the mass of the engine and a_e is its acceleration. An application of the Second Law such as this gives the equation of motion, in this case the equation of motion for the engine.

Similarly, apply the Second Law to the freight car to get its equation of motion,

$$F_{efc} - F_{cfc} = m_{fc}a_{fc}, \tag{1.5}$$

where F_{efc} is the force of the engine on the freight car, F_{cfc} is the force of the caboose on the freight car, m_{fc} is the mass of the freight car, and a_{fc} is its acceleration. Then apply the Second Law to the caboose to get its equation of motion,

$$F_{fcc} = m_c a_c, \tag{1.6}$$

where F_{fcc} is the force of the freight car on the caboose, m_c is the mass of the caboose, and a_c is its acceleration.

Taken together these three equations seem daunting; they certainly do not invite solution. They will, once they are simplified. First, note that the cars of the train are rigidly tied together by the couplers, so they must all have the same acceleration,

$$a_e = a_{fc} = a_c = a. \tag{1.7}$$

Next, notice that by the Third Law the force exerted by the freight car on the engine must be equal and opposite to the force exerted by the engine on the freight car. The directions have been accounted for in the free-body diagrams, so we need only concern ourselves with the magnitudes of the forces. We write $F_{fce} = F_{efc} = F_1$. A similar application of the Third Law shows that the force of the caboose on the freight car must be equal and opposite to the force of the freight car on the caboose. Again, direction has been accounted for in the free-body diagrams, so the magnitudes of the forces become $F_{cfc} = F_{fcc} = F_2$.

With these simplifications we can rewrite the equations of motion. The equation for the engine becomes

$$f_e - F_1 = m_e a, \tag{1.8}$$

that of the freight car becomes

$$F_1 - F_2 = m_{fc}a, \tag{1.9}$$

and that of the caboose becomes

$$F_2 = m_c a. \tag{1.10}$$

So far we are left with three equations in the four unknowns a, F_1, F_2, and f_f. If we can eliminate one of the unknowns then we can solve for the motion of the train. This is good. We know that real trains move, and we are getting close to understanding why.

There are two obvious ways to proceed. One is to measure the acceleration of a real train and calculate the forces involved. The other is to measure the friction and calculate the acceleration and the forces among the cars. We choose to measure the friction.

In many situations the friction is well described by $f_f = \mu N$, where μ is the coefficient of friction and N is the normal force. The coefficient of friction is taken to be a constant of order unity that depends on the nature of the materials that are in contact. The normal force is simply the upward force exerted by the track. In the case of the engine we have shown that it is merely the engine's weight, or $m_e g$, where g is the acceleration of gravity, so $f_f = \mu m_e g$. We can plug this into the equation of motion for the engine, getting

$$\mu m_e g - F_1 = m_e a. \tag{1.11}$$

This presents a strategy for solving for the motion of the train: solve this equation for F_1, plug that into the equation of motion for the freight car and solve for F_2. Plug that into the equation of motion for the caboose and solve for a. The value of a can then be used to fine F_1 and F_2. Proceeding with the plan,

$$F_1 = \mu m_e g - m_e a = m_e(\mu g - a). \tag{1.12}$$

Plugging into the equation of motion for the freight car,

$$\mu m_e g - m_e a - F_2 = m_{fc} a, \tag{1.13}$$

and solving for F_2,

$$F_2 = \mu m_e g - m_e a - m_{fc} a = \mu m_e g - (m_e + m_{fc})a. \tag{1.14}$$

Plugging into the equation of motion for the caboose,

$$\mu m_e g - (m_e + m_{fc})a = m_c a, \tag{1.15}$$

or

$$\mu m_e g = (m_e + m_{fc} + m_c)a = m_{train} a. \tag{1.16}$$

This says that the friction between the engine and the track must be large enough to accelerate the entire train! Amazing! Solving for the acceleration,

$$a = \frac{\mu m_e g}{m_{train}}, \tag{1.17}$$

which says that the acceleration is the friction force divided by the total mass of the train. This makes sense. Note that if there is no freight car or caboose, so their masses are each zero, then

$$a = \frac{\mu m_e g}{m_e} = \mu g, \tag{1.18}$$

which may make sense based on your earlier studies.

Notice how complicated and subtle this analysis is. Let's take a look at what we have accomplished. We have calculated the motion of a train from first principles using Newton's Laws. At the beginning of the seventeenth century nobody was able to do this. By the end of the seventeenth century the most skillful mathematicians and natural philosophers were able to do this. Trains were unknown at the time, so they would have analyzed a horse pulling a wagon, but the analysis would have been the same.

1.2.2 History

What historical change allowed this analytical approach to happen? Math was applied to the description of the physical world. The crucial step before the work of Newton was made by Galileo. He had mentors and spoke with them about what they were doing, but Galileo showed, by laboratory measurement, that an object undergoing constant acceleration has equal velocity changes in equal amounts of time. This is the basis for the definition of constant acceleration familiar to us, $a = \frac{dv}{dt} = const.$, and led to the development of calculus by Newton and Leibniz. Previous to Galileo's measurements and interpretation the weight of opinion was that accelerating objects gained equal amounts of velocity in equal distances, $a = \frac{dv}{dx}$, which we know to be incorrect. This opinion was based on pure thought, not on laboratory measurement. Galileo's discovery is described in the *Dialogues Concerning Two New Sciences* in the section called Day Three.

That was not Galileo's only accomplishment in the *Dialogues*. In the section called Day Two he applied algebra to the description of the bending and failure of beams. Many modern critics have pointed out that his analysis is only partially correct because he did not draw a free-body diagram of the beam and was unclear about the balance of compression and tension at any cross section of a stationary beam. His major accomplishment was the performance of measured experiments and the application of algebra to their results.

1.3 The Equation of Motion for Two Orbiting Bodies

Newton's formulation of gravitation also appeared in the Principia, and we know it as Newton's Law of Universal Gravitation. For two bodies of masses M and m it is written as

$$\vec{F}_g = -\frac{GMm}{r^2}\frac{\vec{r}}{r} = -\frac{GMm}{r^3}\vec{r}, \tag{1.19}$$

where $\vec{F}_g$ is the force of gravity that one of the bodies exerts on the other, G is a constant of nature, $\vec{r}$ is the vector from the body causing the force to the body experiencing it, and r is the magnitude or the vector $\vec{r}$. The negative sign makes the force attractive. We credit Henry Cavendish, 1731-1810, with measuring G, the gravitational constant. He would have claimed that he did something very different - that he determined the mass of the Earth, and from this its density. This result, in itself, is profound, for Cavendish showed that the overall density of the Earth is very similar to that of iron or nickel, suggesting that they are the main constituents of the solid Earth, and, by implication, that the oceans have very limited depth.

Owing to the definition of $\vec{r}$, the force exerted by M on m is equal and opposite to the force exerted by m on M, as required by the Third Law. This will be important in deriving a simple orbit equation.

1.3.1 Choosing a Goal

We would like to know the position as a function of time of orbiting objects, and we would like to derive them from first principles, such as our knowledge of Newton's Laws and Universal Gravitation. This is a challenging goal.

1.3.2 Making Things Complicated

In general, $\vec{F}_g$ is only one of several forces that contribute to the total force acting on an orbiting body. Other forces may include gravitation by other bodies, radiation pressure on the cross-sectional area of the orbiting body, drag owing to the low density of the residual atmosphere, and any other force that can be documented or hypothesized. In general

$$\frac{d\vec{p}}{dt} = \vec{F}_{net}, \tag{1.20}$$

where $\vec{F}_{net}$ is the sum of all acting forces.

1.3.3 Making Things Simple Again

Not wishing to deal with all of this potential complication, let's make some simplifying assumptions and examine the resulting special case.

Assumptions:

Assumption 1. We assume that we have an inertial coordinate system (X, Y, Z) that is adequate for describing the system under study.

An inertial coordinate system is one that is not accelerating. The Earth is spinning on its axis and orbiting the Sun, and the Sun is orbiting the Galaxy, so our location on the surface of the Earth is not an inertial frame. The accelerations must be very small, because we treat the surface of the Earth as an inertial frame all the time.

Problem 2 How nearly inertial is the Earth's reference frame?

Calculate the accelerations of the Earth due to the Earth's rotation, the Earth's orbit about the Sun, and the Sun's orbit about the Milky Way. Assume that all of these motions are circular, and that $a = v^2/r$. Work this problem in mks units and compare the accelerations with that of gravity at the Earth's surface. You will need the radius of the Earth, the radius of the Earth's orbit, and the radius of the Sun's orbit. The first two should be familiar, assume that the third is 8 kiloparsecs and convert to meters. The velocities can be found from the radii and orbital periods. Again, the first two should be familiar, assume that the third is 250,000,000 years. In converting to seconds, show that the number of seconds in a year is very nearly $\pi \times 10^7$.

Following Figure 1.2-1 in the text, we number the masses under study in this system from 1 to n, so their masses are

$$m_1, m_2, ...m_i, m_j, m_k, ...m_n \tag{1.21}$$

and their position vectors are

$$\vec{r}_1, \vec{r}_2, ...\vec{r}_i, \vec{r}_j, \vec{r}_k, ...\vec{r}_n. \tag{1.22}$$

Problem 3 You should convince yourself that the vector from m_n to m_i, called $\vec{r}_{ni}$ in the text, is

$$\vec{r}_{ni} = \vec{r}_i - \vec{r}_n. \tag{1.23}$$

You should do this in three ways: by setting $\vec{r}_i = 0$, by setting $\vec{r}_n = 0$, and by making a sketch of the vectors in (X, Y, Z) space.

I am somewhat displeased with the notation used in the text, because the subscript n carries a special meaning attached to the last named mass. I prefer replacing n with j, which has no special meaning, to make the statement more general. Thus, I would write

$$\vec{r}_{ji} = \vec{r}_i - \vec{r}_j. \tag{1.24}$$

Continuing with the general notation, this means that the force exerted on mass m_i by mass m_j is

$$\vec{F}_{gji} = -\frac{Gm_i m_j}{r_{ji}^3}\vec{r}_{ji}. \tag{1.25}$$

Here I have changed the notation slightly. The text would call this force $\vec{F}_{gj}$, I have added the additional subscript i for clarity. The sum of all such gravitational forces acting on mass m_i by all possible masses m_j is then

$$\vec{F}_{gi} = -\frac{Gm_i m_1}{r_{1i}^3}\vec{r}_{1i} - \frac{Gm_1 m_2}{r_{2i}^3}\vec{r}_{2i} - ... - \frac{Gm_i n_n}{r_{ni}^3}\vec{r}_{ni} = -Gm_i \sum_{j=1}^{n} \frac{m_j}{r_{ji}^3}\vec{r}_{ji}, \tag{1.26}$$

which partially returns us to the notation of the book. Here we face a small problem: what do we do when $j = i$? Can a body exert a force on itself? If it can, this seems to imply that the body can be subdivided into two interacting parts, in which case those parts must exert equal and opposite forces by the Third Law. If so, the forces cancel in the sum. Thus, we take as given that a body cannot exert a net nonzero gravitational force on itself until we can be convinced otherwise. The total gravitational force on mass i becomes

$$\vec{F}_{gi} = -Gm_i \sum_{j \neq i}^{n} \frac{m_j}{r_{ji}^3} \vec{r}_{ji}. \tag{1.27}$$

We must allow for the possibility that gravity is not the only force, so that $\vec{F}_{tot} = \vec{F}_{grav} + \vec{F}_{other}$, where the other force is yet to be specified. This total force then equals the time rate of change of momentum.

Assumption 2. We assume that only two bodies are present in this space that is otherwise empty, or so nearly empty that emptiness may be assumed. Thus, $i = 1, 2$.

Assumption 3. We assume that the bodies are spherically symmetric so that their gravitation is mathematically identical to that of point masses located at their centers. This assumption makes gravitational forces central forces. To say this another way, the gravitational forces are then antiparallel to the radius vectors, so any cross product, like that for torque, $\vec{\tau} = \vec{r}_i \times \vec{F}_g$, must be zero. This means that gravity cannot cause torques. We will have to relax this assumption in Chapter 3 when we consider the Earth's equatorial bulge, which has a non-spherical mass distribution.

Assumption 4. We assume that the bodies have constant masses, so

$$\frac{d\vec{p}_i}{dt} = m_i \frac{d\vec{v}_i}{dt} = m_i \frac{d\dot{\vec{r}}_i}{dt} = m_i \ddot{\vec{r}}_i. \tag{1.28}$$

This turns out to be a greater help to us than might originally be imagined. If the masses are constant then the number of time-dependent variables that we wish to solve for is reduced from eight to six, that is, from one mass and three positions for each object to just the three positions for each object.

Let $\vec{r}$ be the vector from M to m. Following the notation in the text we will let body 1 have mass M, so $m_1 = M$, and let body 2 have mass m, so $m_2 = m$. Let the position of M be $\vec{r}_1$ and that of m be $\vec{r}_2$. The position of m relative to M is then $\vec{r}_{21} = \vec{r}_1 - \vec{r}_2 = \vec{r}$, and the position of M relative to m is $\vec{r}_{12} = \vec{r}_2 - \vec{r}_1 = -\vec{r}$. These vectors can be used to calculate the gravitational forces.

The gravitational force exerted by M on m is

$$m\ddot{\vec{r}}_2 = -\frac{GMm}{r^2} \frac{\vec{r}}{r}, \tag{1.29}$$

and the gravitational force exerted by m on M is

$$M\ddot{\vec{r}}_1 = \frac{GMm}{r^2}\frac{\vec{r}}{r}. \tag{1.30}$$

The corresponding accelerations are

$$\ddot{\vec{r}}_2 = -\frac{GM}{r^3}\vec{r}, \tag{1.31}$$

and

$$\ddot{\vec{r}}_1 = \frac{Gm}{r^3}\vec{r}. \tag{1.32}$$

Subtracting the second acceleration from the first gives

$$\ddot{\vec{r}}_2 - \ddot{\vec{r}}_1 = \ddot{\vec{r}} = -\frac{G(M+m)}{r^3}\vec{r}. \tag{1.33}$$

This equation involves only the relative position, $\vec{r}$, and converts the problem of gravitational motion into an equivalent one-body problem, with a body whose mass is the sum of the two masses. We now have only three unknowns, the three vector components of $\vec{r}$, which is a great simplification.

It is customary to define $\mu = G(M+m)$, and write

$$\ddot{\vec{r}} + \frac{\mu}{r^3}\vec{r} = 0. \tag{1.34}$$

This is the equation that will form the basis of our work for the remainder of the semester. In nearly all the cases that we will study M is much greater than m, so $\mu \sim GM$.

Assumption 5. We assume that the bodies have distinctly different masses, with the more massive body having mass M and the less massive body having mass m, such that $M \gg m$. This assumption sets us up for cases like planetary motion, in which M is the mass of the Sun and m is the mass of a planet, or satellite motion, in which M is the mass of the Earth and m is the mass of the satellite. In either case $M \gg m$ is a very good assumption. If this assumption breaks down then we will have to modify the subsequent equation. The assumption does not apply to binary star systems in which the two stars have similar masses.

1.4 Partial Solutions of the Equation of Motion

1.4.1 Constants of the Motion

Specific Mechanical Energy

Our goal is a solution to the equation of motion, that is, the functional form for $\vec{r}(t)$. This is a lofty goal that will require a lot of work. It turns out to be relatively easy to learn a lot about $\vec{r}(t)$ without actually finding the solution, and that is how we will start.

To someone experienced in solving the equation of motion it is natural to multiply both sides by $\dot{\vec{r}}$ and integrate. This may look like a trick at first, but it is not, for many equations of motion can be integrated in this way. Experience shows that it is an obvious thing to do with an obvious result. Dot multiply on the left by $\dot{\vec{r}}$ to get

$$\dot{\vec{r}} \cdot \ddot{\vec{r}} + \dot{\vec{r}} \cdot \frac{\mu}{r^3}\vec{r} = 0. \tag{1.35}$$

We know that $\dot{\vec{r}} = \vec{v}$, and we can show that $\vec{a} \cdot \dot{\vec{a}} = a\dot{a}$. Rather than assign this as homework, here is the proof. We know

$$\vec{a} \cdot \vec{a} = a^2, \tag{1.36}$$

so

$$\frac{d}{dt}(\vec{a} \cdot \vec{a}) = \frac{d}{dt}a^2, \tag{1.37}$$

and

$$2\vec{a} \cdot \dot{\vec{a}} = 2a\dot{a}, \tag{1.38}$$

so

$$\vec{a} \cdot \dot{\vec{a}} = a\dot{a}. \tag{1.39}$$

Now our equation becomes

$$\vec{v} \cdot \dot{\vec{v}} + \frac{\mu}{r^3}\dot{\vec{r}} \cdot \vec{r} = 0, \tag{1.40}$$

giving

$$v\dot{v} + \frac{\mu}{r^3}r\dot{r} = 0. \tag{1.41}$$

This can be integrated by inspection, because

$$\frac{d}{dt}\frac{v^2}{2} = v\dot{v}, \tag{1.42}$$

and

$$\frac{d}{dt}\left(\frac{-\mu}{r}\right) = \frac{\mu}{r^2}\dot{r}. \tag{1.43}$$

Try these for yourself. This gives

$$\frac{d}{dt}\left(\frac{v^2}{2} - \frac{\mu}{r}\right) = 0. \tag{1.44}$$

This result can be made even more general by noticing that

$$\frac{d}{dt}\left(\frac{v^2}{2} + c - \frac{\mu}{r}\right) = 0 \tag{1.45}$$

where c is an arbitrary constant. Since the net time derivative is zero we may take the function in parentheses to be a constant of time and write

$$\mathcal{E} = \frac{v^2}{2} + \left(c - \frac{\mu}{r}\right) \tag{1.46}$$

where $\mathcal{E}$ is a constant. The first term in $\mathcal{E}$ is the kinetic energy per unit mass, or specific kinetic energy. The second term is the gravitational potential energy per unit mass (or the specific potential) plus an arbitrary constant. If we choose the constant to be zero then the specific potential goes to zero from below as r becomes large. Choosing $c = 0$ leaves us with the specific mechanical energy,

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r}, \tag{1.47}$$

which is a constant of the motion. Another way of saying this is that the specific mechanical energy is conserved.

Specific Angular Momentum

Experience shows that the dot multiplication above is an obvious thing to do and reliably leads to the conservation of energy. It is less obvious to try cross multiplication by $\vec{r}$, but it can do no harm, and will eliminate one term from the equation of motion because any vector crossed with itself gives zero. Try

$$\vec{r} \times \ddot{\vec{r}} + \vec{r} \times \frac{\mu}{r^3}\vec{r} = 0. \tag{1.48}$$

The second term must be zero, leaving

$$\vec{r} \times \ddot{\vec{r}} = 0. \tag{1.49}$$

This, too, can be integrated by inspection by trying

$$\frac{d}{dt}\left(\vec{r} \times \dot{\vec{r}}\right) = \dot{\vec{r}} \times \dot{\vec{r}} + \vec{r} \times \ddot{\vec{r}}. \tag{1.50}$$

Any vector crossed with itself must again give zero, leaving the other term, so

$$\frac{d}{dt}\left(\vec{r} \times \dot{\vec{r}}\right) = \frac{d}{dt}\left(\vec{r} \times \vec{v}\right) = 0. \tag{1.51}$$

Thus, we have discovered another constant of the motion, $\vec{r} \times \vec{v}$. This is just the specific angular momentum, and we write

$$\vec{h} = \vec{r} \times \vec{v}. \tag{1.52}$$

This result turns out to be more subtle than the previous one, for $\vec{h}$ must be constant in both magnitude and direction, forcing the vectors $\vec{r}$ and $\vec{v}$ to define a plane, since $\vec{h}$ must be perpendicular to both $\vec{r}$ and $\vec{v}$ by the definition of the cross product. Thus, the orbital motion of the two bodies is confined to a plane when specific angular momentum is conserved.

Writing out the magnitude of the cross product gives

$$h = rv \sin\gamma = rv \cos\phi, \tag{1.53}$$

where γ is the angle between $\vec{r}$ and $\vec{v}$ when they are drawn tail to tail, and ϕ is the complement of γ. Please refer to Figure 1.4-1 in the text. γ is the angle between the local vertical and $\vec{v}$, and is called the zenith angle. Thus, ϕ is the angle between the horizontal and $\vec{v}$, and is called the flight-path angle. Of these two angles ϕ is usually the more easily observable, and will appear in future chapters, so it it convenient to introduce it here. The sign of ϕ is the same as the sign of $\vec{r} \cdot \vec{v}$.

1.4.2 The Trajectory Equation

The next manipulation of the equation of motion is brilliant, very subtle, and not at all obvious. It was first performed by Newton, and I wonder how he did it so long ago, before our current vector notation had been developed. Write the equation of motion as

$$\ddot{\vec{r}} = -\frac{\mu}{r^3}\vec{r}, \tag{1.54}$$

and cross multiply by $\vec{h}$ from the right, while at the same time changing order of operation and sign on the right-hand side of the equation, to get

$$\ddot{\vec{r}} \times \vec{h} = \frac{\mu}{r^3}\big(\vec{h} \times \vec{r}\big). \tag{1.55}$$

Problem 4 The time derivative of the cross product

Compare the left-hand side of the equation with

$$\frac{d}{dt}\big(\dot{\vec{r}} \times \vec{h}\big) \tag{1.56}$$

to show that they are equal.

The right-hand side can also be expressed as a time derivative. Expand $\vec{h}$ to get

$$\begin{aligned}\frac{\mu}{r^3}\big(\vec{h} \times \vec{r}\big) = \frac{\mu}{r^3}\big(\vec{r} \times \vec{v}\big) \times \vec{r} = \frac{\mu}{r^3}\Big(\vec{v}\big(\vec{r}\cdot\vec{r}\big) - \vec{r}\big(\vec{r}\cdot\vec{v}\big)\Big) \\ = \frac{\mu}{r}\vec{v} - \frac{\mu\dot{r}}{r^2}\vec{r}.\end{aligned} \tag{1.57}$$

The next-to-last step makes use of a vector identity that appears in Appendix C of the text. There are two closely-related identities, the first of which I know as the $BAC - CAB$ rule, the name providing an aid to memory:

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A}\cdot\vec{C}) - \vec{C}(\vec{A}\cdot\vec{B}), \tag{1.58}$$

and

$$(\vec{A} \times \vec{B}) \times \vec{C} = \vec{B}(\vec{A}\cdot\vec{C}) - \vec{A}(\vec{B}\cdot\vec{C}). \tag{1.59}$$

We note that

$$\mu\frac{d}{dt}\left(\frac{\vec{r}}{r}\right) = \frac{\mu}{r}\vec{v} - \frac{\mu\dot{r}}{r^2}\vec{r}. \tag{1.60}$$

Thus, we can write

$$\frac{d}{dt}\left(\dot{\vec{r}} \times \vec{h}\right) = \mu \frac{d}{dt}\left(\frac{\vec{r}}{r}\right). \tag{1.61}$$

Integrating both sides gives

$$\dot{\vec{r}} \times \vec{h} = \mu \frac{\vec{r}}{r} + \vec{B}, \tag{1.62}$$

where $\vec{B}$ is a vector constant of integration, and turns out to be an additional constant of the motion. Dot multiplying by $\vec{r}$ on the left gives

$$\vec{r} \cdot \dot{\vec{r}} \times \vec{h} = \vec{r} \cdot \mu \frac{\vec{r}}{r} + \vec{r} \cdot \vec{B}. \tag{1.63}$$

There is a vector identity

$$\vec{a} \cdot \vec{b} \times \vec{c} = \vec{a} \times \vec{b} \cdot \vec{c}, \tag{1.64}$$

and we already know $\vec{a} \cdot \vec{a} = a^2$, so

$$\vec{r} \times \dot{\vec{r}} \cdot \vec{h} = \mu r + \vec{r} \cdot \vec{B}, \tag{1.65}$$

giving the scalar equation

$$h^2 = \mu r + rB \cos \nu, \tag{1.66}$$

where ν is the angle between B and r. Solving for r gives

$$r = \frac{h^2/\mu}{1 + (B/\mu)\cos \nu}. \tag{1.67}$$

Very nice, but what has this accomplished?

1.5 Conic Sections

Conic sections are formed by the intersection of a cone and a plane, as in Figure 1.5-2 in the text. A true mathematical cone looks like two ice cream cones set parallel with points touching. We will see that conic sections include circles, ellipses, parabolas, hyperbolas, lines, and a point. Conic sections were known and studied by the Greeks. Knowledge of conic sections would have been common among educated people in the time of Kepler and Newton.

The following derivations draw heavily on material from Wolfram MathWorld, cited as:

Weisstein, Eric W. “Ellipse.” From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/Ellipse.html

Weisstein, Eric W. “Hyperbola.” From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.c/Hyperbola.html

and

Weisstein, Eric W. “Conic Section Directrix.” From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/ConicSectionDirectrix.html

Additional valuable material on ellipses may be found at

http://www.oc.nps.navy.mil/g̃arfield/ellipse_app2.pdf

1.5.1 Polar Equations of Conics

The equation of an ellipse with the coordinate origin located at its center is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{1.68}$$

where a is the semi-major axis and b is the semi-minor axis. This is probably familiar from your earlier studies. This is not the only way to describe an ellipse, and, in fact, we seek a description in polar coordinates with the origin at a focus of the ellipse. The motivation for this is a desired comparison with the trajectory equation. Remember Kepler’s First Law? “The orbit of each planet is an ellipse with the Sun at one focus.” If the Sun is the center of the Solar System then its center is the logical place to put the origin of a set of polar coordinates, so we wish to move the coordinate origin to a focus. From the diagram we see that

$$x = c + r\cos\theta,\ y = r\sin\theta. \tag{1.69}$$

Making the obvious substitutions and multiplying both sides by a^2b^2 gives

$$b^2c^2 + 2b^2cr\cos\theta + b^2r^2\cos^2\theta + a^2r^2\sin^2\theta = a^2b^2. \tag{1.70}$$

We are going to manipulate this equation and compare the result with the trajectory equation. We will see that the two equations are the same. The web sites listed above show that there are several parameters other than a and b that are used to describe ellipses. Two common ones are the eccentricity, e, and the half-focal separation, c, which is the distance from the center of the ellipse to either focus. Mathematically,

$$e = \sqrt{1 - \frac{b^2}{a^2}}, \tag{1.71}$$

and

$$c = ae. \tag{1.72}$$

We use these expressions to eliminate b and c by replacing them with terms in a and e, using

$$b^2 = a^2(1 - e^2) \tag{1.73}$$

and the trigonometric identity

$$\sin^2\theta = 1 - \cos^2\theta \tag{1.74}$$

to get

$$\begin{aligned} a^2(1-e^2)a^2e^2 + a^2r^2 + 2a^2(1-e^2)aer\cos\theta \\ + a^2(1-e^2)r^2\cos^2\theta - a^2r^2\cos^2\theta \\ = a^2a^2(1-e^2). \end{aligned} \tag{1.75}$$

Dividing through by a^2 is an obvious step. Regrouping gives

$$r^2 + a^2(1-e^2)(e^2-1) + 2a(1-e^2)er\cos\theta - e^2r^2\cos^2\theta = 0. \tag{1.76}$$

Next multiply through by -1 and isolate the r^2 term to get

$$r^2 = a^2(1-e^2)^2 - 2a(1-e^2)er\cos\theta + e^2r^2\cos^2\theta. \tag{1.77}$$

Taking the square root gives

$$r = \pm\big(er\cos\theta - a(1-e^2)\big), \tag{1.78}$$

where the sign is to be determined. The radius and a must be positive, and e may be positive or zero. Taking $e = 0$ gives

$$r = \pm(-a), \tag{1.79}$$

so we must choose the leading negative sign, giving

$$r = a(1 - e^2) - er\cos\theta. \tag{1.80}$$

Grouping the terms in r gives

$$r\big(1 + e\cos\theta\big) = a(1 - e^2), \tag{1.81}$$

and solving for r gives

$$r = \frac{a(1 - e^2)}{1 + e\cos\theta}. \tag{1.82}$$

This equation is commonly written as

$$r = \frac{p}{1 + e\cos\theta} \tag{1.83}$$

where p is called the semi-latus rectum, or parameter, and e is the eccentricity. Thus,

$$p = a(1 - e^2). \tag{1.84}$$

Compare this result with the trajectory equation

$$r = \frac{h^2/\mu}{1 + (B/\mu)\cos\theta}. \tag{1.85}$$

We identify

$$p = \frac{h^2}{\mu},\ e = \frac{B}{\mu},\ \nu = \theta. \tag{1.86}$$

Thus, we have shown that ellipses match the mathematical form of gravitational orbits.

This is not the only possibility: in general gravitational orbits are conic sections, which include circles, ellipses, parabolas, and hyperbolas. This is most easily discovered by examining the directrix definition of conic sections. The conic sections may be defined using a line, called a directrix, and a point, called a focus. The conic section itself is "the locus of points whose distance from the focus is proportional to the horizontal distance from the directrix," *given a vertical directrix.* (The quote is from Weisstein's article "Conic Section Directrix", reference at the top of this section. The italic emphasis is mine.) If the two distances are equal then the conic section is a parabola. If the distance to the focus is smaller than the horizontal distance to the directrix then the conic is an ellipse. If the distance to the focus is larger than the horizontal distance to the directrix then the conic section is a hyperbola.

It is tempting to geuss that when the distance to the focus equals the distance to the directrix the result is a parabola. Let's find out. Consider a focus at position $(a, 0)$ and a directrix at $x = -a$, and a general point (x, y). The horizontal distance to the directrix is $x+a$ and the distance to the focus is $\sqrt{(x-a)^2 + y^2}$. Equate and square both sides to get

$$x^2 + 2ax + a^2 = x^2 - 2ax + a^2 + y^2. \tag{1.87}$$

Subtracting x^2 and a^2 from both sides leaves

$$y^2 = 4ax, \tag{1.88}$$

which is the familiar equation of a parabola.

Switching to polar coordinates with origin at the focus gives

$$x = a - r\cos\theta, y = r\sin\theta, \tag{1.89}$$

so the equation of the parabola becomes

$$r^2 \sin^2\theta = 4a^2 - 4ar\cos\theta. \tag{1.90}$$

Substituting $1 - \cos^2\theta$ for the $\sin^2\theta$ term and reorganizing gives

$$r^2 = (2a - r\cos\theta)^2. \tag{1.91}$$

Taking the square root gives

$$r = \pm(2a - r\cos\theta). \tag{1.92}$$

We again require that r be positive, and examine the result when $\theta = 0$ and $r = a$, giving

$$a = \pm(2a - a), \tag{1.93}$$

where the choice of the positive sign is clear. Then

$$r = 2a - r\cos\theta, \tag{1.94}$$

and solving for r gives

$$r = \frac{2a}{1 + \cos\theta}, \tag{1.95}$$

which matches the trajectory equation, with $p = 2a$ and $e = 1$. Thus, parabolas also match the trajectory equation.

Rather than show that a hyperbola also matches the trajectory equation we choose to work the general result for any conic. Start with Cartesian coordinates (x', y') with an origin $\mathcal{O}'$. Place a focus at $(c, 0)$ and a vertical directrix at $(c + d, 0)$. This can describe a general case if we allow d to be positive or negative. Choose the origin of a second coordinate system at the focus. Label the coordinates relative to this origin (x, y) and call the origin $\mathcal{O}$. Relative to this origin the focus is at $(0, 0)$, the directrix is at $x = d$, and an arbitrary point has coordinates (x, y). Clearly, the distance of this arbitrary point from the origin is $\sqrt{x^2 + y^2}$ and its horizontal distance from the directrix is $(d - x)$. We seek the equation in (x, y) coordinates that defines figures that have a constant ratio of proportionality, e, between these distances, so, working with the squares to get positive-definite quantities

$$x^2 + y^2 = e^2(d - x)^2 = e^2(x^2 - 2xd + d^2), \tag{1.96}$$

or,

$$y^2 = (e^2 - 1)x^2 - 2e^2dx + e^2d^2. \tag{1.97}$$

We may also define polar coordinates (r, θ) with respect to origin $\mathcal{O}$ such that

$$x = rcos\theta, y = r\sin\theta, \sin^2\theta = 1 - \cos^2\theta, \tag{1.98}$$

giving

$$r^2(1 - cos^2\theta) = (e^2 - 1)r^2\cos^2\theta - 2e^2dr\cos\theta + e^2d^2. \tag{1.99}$$

Recognize that both sides of the above equation have a term $-r^2\cos^2\theta$ that cancels, and reorganize to get

$$r^2 = e^2(r^2\cos^2\theta - 2rd\cos\theta + d^2) = e^2(r\cos\theta - d)^2. \tag{1.100}$$

Taking the square root gives

$$r = \pm e(r\cos\theta - d). \tag{1.101}$$

When $\theta = 90^o$ we obtain

$$r = \pm e(-d), \tag{1.102}$$

which leads us to choose the leading negative sign, leaving

$$r = e(d - r\cos\theta). \tag{1.103}$$

Solving for r gives the desired result,

$$r = \frac{ed}{1 + e\cos\theta}. \tag{1.104}$$

Problem 5 The polar form of the hyperbola

Starting with the equation for a hyperbola with its symmetry point at the origin, its foci at $(c, 0)$ and $(-c, 0)$, and its vertices at $(a, 0)$ and $(-a, 0)$,

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{1.105}$$

put the equation in the form of the trajectory equation with the origin at the focus $(c, 0)$. You will need $c = ea$ and $e = \sqrt{1 + \frac{b^2}{a^2}}$. Notice the sign change in the formula for e compared with that for the ellipse.

The conic sections are distinguished by their eccentricities. The circle and parabola are special cases, with $e = 0$ and $e = 1$, respectively. Ellipses have $0 < e < 1$ and hyperbolas have $e > 1$. The conic sections are the only possible paths for two-body orbits.

1.6 Properties of Conic-Section Orbits

As mentioned above, conic sections are the only possible orbits for a two-body system. In real systems, e.g., the Solar System, deviations from conic sections are caused by the presence of additional bodies.

We can now understand Kepler's First Law: "The orbit of each planet is an ellipse with the Sun at one focus." This is a consequence of gravitationally bound, enduring, two-body orbits, and the fact that the Sun is much more massive than any of the planets. Parabolic and hyperbolic orbits are allowed, but they do not endure because those figures extend to infinity. Actually, the center of mass of the two bodies is at the focus, but the Sun is so massive that the center of mass is very close to the center of the Sun. We have seen that the elliptical orbit takes place in a plane that is fixed in space, and that r and v change in such a way that $\mathcal{E}$, $\vec{h}$, and $\vec{B}$ remain constants of the motion.

A general conic section has two foci. The circle and parabola are special cases - the two foci are coincident in the circle and one focus is at infinity for the parabola. The empty focus appears to have no particular importance in orbital mechanics. Similarly, the directrix has no particular importance. The height of the orbit at the location of the focus is the parameter or semi-latus rectum, p. The length of the orbit along the line defined by the foci is called the major axis and has extent $2a$, giving the semi-major axis length a. In a circle $2a$ is the diameter and a is the radius. For the parabola $2a$ is infinite and for the hyperbola it is taken as negative. The distance between the foci is $2c$. For the circle c is zero, for the ellipse it is positive, for the parabola infinite, and for the hyperbola negative. Please refer to figure 1.5-3 in the text. For all conics except the parabola

$$p = a(1 - e^2) \tag{1.106}$$

and

$$e = \frac{c}{a}. \tag{1.107}$$

Note that for a parabola $e = 1$ but $p \neq 0$.

The extreme points of the orbit are known as turning points or apses, singular apse, from the Greek via Latin for arch. The point on the orbit nearest the occupied focus is called the periapsis and the point farthest from the occupied focus is called the apoapsis. The names can change according to the gravitating body at the focus: perigee and apogee for the Earth, perihelion and aphelion (pronounced afelion) for the Sun, periselenium and aposelenium for the Moon (don't get caught off guard by the terms of mixed origin, perilune and apolune), and perigalacticon and apogalacticon for galaxies. Notice that these terms are not uniquely defined for a circle, and that the apoapsis has no definite meaning for a parabola or hyperbola.

We can calculate the distance to the periapsis and apoapsis by setting $\nu = 0^o$ and $\nu = 180^o$ in the trajectory equation, getting

$$r_{min} = r_{peri} = r_p = \frac{p}{1+e} = a\frac{1-e^2}{1+e} = a(1-e), \tag{1.108}$$

and

$$r_{max} = r_{apo} = r_a = \frac{p}{1-e} = a\frac{1-e^2}{1-e} = a(1+e). \tag{1.109}$$

1.6.1 Relating the Constants of the Motion to the Geometry of the Orbit

The constants of motion are $\mathcal{E}$, $\vec{h}$, and $\vec{B}$. The geometry of the orbit is described by p, a, e, and ν. The quantities p, a, and e are not independent - any two can be used to find the third because $p = a(1 - e^2)$. Comparing the trajectory equation with the general polar equation for conic section orbits shows that $p = h^2/\mu$, $e = B/\mu$, and that the reference direction for measuring ν is provided by the direction of $\vec{B}$.

h Determines p - Newton's Cannon

Of the constants of motion, the quantity h alone determines p. This is obvious from the comparison above, which gives $p = h^2/\mu$. This is also demonstrated by a thought experiment proposed by Newton, and illustrated in Figure 1.6-1 in the text. Imagine a cannon on a mountain top. The height of the mountain determines $\vec{r}$ at the moment that teh shot is fired. The cannon can be fired successively with greater and greater charges of powder. Assume a limitless supply of cannonballs and powder, and ignore air resistance. Aim the cannon along the local horizontal, so that the flight-path angle, ϕ, is zero. Shots using increasing amounts of powder will propel the cannonball in ever broader arcs. More powder increases v, which increases h, which increases p.

$\mathcal{E}$ Determines a

Of the constants of the motion, the quantity $\mathcal{E}$ alone determines a. We can show this by considering periapsis and apoapsis of conic-section orbits. At these places the velocity is tangent to the orbit and perpendicular to the radius. Another way of saying this is that the flight-path angle, ϕ, is zero. Then

$$h = r_p v_p = r_a v_a, \tag{1.110}$$

and

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} = \frac{h^2}{2r_p^2} - \frac{\mu}{r_p}, \tag{1.111}$$

but

$$r_p = a(1 - e) \tag{1.112}$$

and

$$h^2 = p\mu = \mu a(1 - e^2) \tag{1.113}$$

so

$$\begin{aligned}\mathcal{E} = \frac{\mu a(1-e^2)}{2a^2(1-e)^2} - \frac{\mu}{a(1-e)} &= \frac{\mu}{2a}\frac{(1+e)}{(1-e)} - \frac{\mu}{a}\frac{1}{(1-e)} \\ &= \frac{\mu}{2a}\frac{1}{(1-e)}(1+e-2) = \frac{\mu}{2a}\frac{(e-1)}{(1-e)} = -\frac{\mu}{2a}. \quad (1.114)\end{aligned}$$

Thus, $\mathcal{E}$ determines a.

p and a, or $\mathcal{E}$ and h, Determine e

We already know

$$p = a(1-e^2). \tag{1.115}$$

Solving for e gives

$$e = \sqrt{1 - p/a}, \tag{1.116}$$

but

$$p = h^2/\mu, \; a = -\mu/2\mathcal{E}, \tag{1.117}$$

which can be used to get the desired result

$$e = \sqrt{1 + \frac{2\mathcal{E}h^2}{\mu^2}}. \tag{1.118}$$

1.6.2 Some Important Properties of Individual Conic Orbits

Elliptical Orbits and Kepler's Laws

We know $\vec{h} = \vec{r} \times \vec{v}$. The magnitude of $\vec{h}$ is the magnitude of $\vec{r}$ times the tangential component of $\vec{v}$, which is $r\dot{\nu}$. This is illustrated in Figures 1.7-2 and 1.7-3 in the text. Thus

$$h = r^2 \frac{d\nu}{dt} \tag{1.119}$$

and

$$dt = \frac{r^2}{h} d\nu. \tag{1.120}$$

But the area swept out by the radius vector is

$$dA = \frac{1}{2} r^2 d\nu, \tag{1.121}$$

so

$$d\nu = \frac{2}{r^2} dA, \tag{1.122}$$

and

$$dt = \frac{r^2}{h} \frac{2}{r^2} dA = \frac{2}{h} dA. \tag{1.123}$$

This is Kepler's Second Law - the line joining any planet to the Sun sweeps out equal areas in equal times. It is a consequence of conservation of angular momentum.

If we integrate both sides of the above equation over one complete elliptical orbit we get $A = \pi ab$ and

$$T_p = \frac{2\pi ab}{h}, \tag{1.124}$$

where T_p is the orbit's period, but

$$b = \sqrt{a^2 - c^2} = \sqrt{a^2 - a^2 e^2} = \sqrt{a^2(1 - e^2)} = \sqrt{ap}. \tag{1.125}$$

We already know that $h = \sqrt{\mu p}$, so

$$T_p = 2\pi a \frac{\sqrt{ap}}{\sqrt{\mu p}} = \frac{2\pi}{\sqrt{\mu}} a^{3/2}, \tag{1.126}$$

or

$$T_p^2 = \frac{4\pi^2}{\mu} a^3. \tag{1.127}$$

This is Kepler's Third Law - the squares of the periods of any two planets are in the same proportion as the cubes of their mean distances from the Sun. We have shown that Kepler's Laws are a consequence of the conic nature of gravitational orbits. This is quite an accomplishment!

Problem 6 The mass of the Sun

Reminding ourselves that when the mass of the gravitating body is much greater than the mass of the orbiting body, as in the case of planetary motion in the Solar System, $\mu = GM$. In such a case Kepler's Third Law can be solved for M, giving a method to calculate the mass of the Sun. Use the known values of T and a for the Earth to calculate M.

Circular Orbits

Circles are a special case of ellipses in which $a = b = r$. We define the circular speed, v_{cs}, the circular radius, r_{cs}, and the orbital period at the circular speed, T_{cs}. Then we can find the circular speed in two ways. First,

$$T_{cs} = \frac{2\pi}{\sqrt{\mu}} r_{cs}^{3/2} = \frac{2\pi r_{cs}}{v_{cs}}, \tag{1.128}$$

from the circumference divided by the speed, or

$$\mathcal{E} = -\frac{\mu}{2r_{cs}} = \frac{v_{cs}^2}{2} - \frac{\mu}{r_{cs}}, \tag{1.129}$$

from the energy equation. Either can be solved to give

$$v_{cs} = \sqrt{\frac{\mu}{r_{cs}}}. \tag{1.130}$$

Note the importance of Problem 1.17 in the text and Newton's cannon to the comparison of circular and elliptical orbits.

Parabolic Orbits

Parabolic orbits have two useful properties. First, a parabola's radius at periapsis is especially simple, as shown in Figure 1.9-1 of the text. Because $e = 1$ the orbit in focal polar coordinates is

$$r = \frac{p}{1 + \cos\nu}. \tag{1.131}$$

Periapsis occurs when $\nu = 0^o$, where $\cos\nu = 1$, so

$$r = r_{peri} = r_p = \frac{p}{2}. \tag{1.132}$$

The parabolic orbit has zero specific mechanical energy, making it easy to calculate the escape speed,

$$\frac{v_{esc}^2}{2} - \frac{\mu}{r} = 0, \tag{1.133}$$

so

$$v_{esc} = \sqrt{\frac{2\mu}{r}}. \tag{1.134}$$

At escape v_{esc} goes to zero and r goes to infinity, which agrees with $\mathcal{E} = 0$.

Hyperbolic Orbits

An important property of hyperbolic orbits comes from the fact that the incoming and outgoing parts of the orbit have a pair of straight-line asymptotes that are separated by the turning angle, δ, which can be calculated because $\sin\frac{\delta}{2} = \frac{1}{e}$. This is shown in Figure 1.10-1 of the text.

The hyperbolic orbit has positive specific energy, so at any distance r from the force center the object in hyperbolic orbit is moving faster than an object in a parabolic orbit. This gives rise to the hyperbolic excess speed. Imagine putting an object into hyperbolic orbit by burning a rocket engine. We compare the burnout point with infinity using the energy equation,

$$\mathcal{E} = \frac{v_{bo}^2}{2} - \frac{\mu}{r_{bo}} = \frac{v_\infty^2}{2} - \frac{\mu}{r_\infty}, \tag{1.135}$$

which gives

$$v_\infty^2 = v_{bo}^2 - \frac{2\mu}{r_{bo}} = v_{bo}^2 - v_{esc}^2. \tag{1.136}$$

This is the hyperbolic excess speed.

1.6.3 Relationships Among the Conics

We really have two general types of orbits - elliptical and hyperbolic - and two special cases - circular and parabolic. A parabolic orbit is a special intermediate case between elliptical and hyperbolic. A circular orbit is the limiting case of an ellipse when b approaches a. Let's compare the two general cases, ellipses and hyperbolas.

An ellipse can be drawn using two foci separated by a distance of $2c$ connected by a piece of string of length $2a$ with $a > c$. Put a pencil in the string, keep both segments of string tight, and move the pencil to draw an ellipse. When the pencil is on the line bisecting the foci and perpendicular to the line joining them then it defines two right triangles, each of base c, height b, and hypotenuse a, so

$$a^2 = b^2 + c^2. \tag{1.137}$$

The arms of hyperbolas are limited by asymptotes that can be used to define right triangles. The foci of the two branches of a hyperbola are separated by a distance of $2c$, as for the ellipse. The apexes are the two branches are separated by $2a$, as are the apses of the ellipse, but the foci of the hyperbola are both outside the apexes, so $c > a$, and the situation is reversed relative to that of the ellipse. An arc of a circle of radius c with its center at the intersection of the asymptotes intersects the asymptotes at height b, making a right triangle with base a and hypotenuse c. Thus,

$$c^2 = a^2 + b^2. \tag{1.138}$$

The roles of c and a have interchanged in going from the ellipse to the hyperbola.

A Reminder: Zenith and Flight-Path Angles

The angle between the velocity vector, $\vec{v}$, and the local vertical is called the zenith angle, γ. The angle between the velocity vector and the local horizontal is the flight-path angle, ϕ. The two angles are complements.

More on Newton's Cannon

Consider, again, Newton's cannon on a mountain top. Let it be supplied with an inexhaustible supply of powder, cannonballs, and all other necessary equipment. Aim the cannon horizontally, so that the zenith angle is 90^o and the flight-path angle is zero. The magnitude of the angular momentum is $h = rv \sin\gamma = rv \cos\phi = rv$. There are sub-orbital trajectories for the cannonball that are ellipses with the center of the Earth being at the focus farther from the cannon, so the cannonball starts at apoapsis (or apogee). These orbits may be thought of as ballistic missile trajectories, for they intersect the Earth. They are mostly interior to the circular orbit. We imagine a first shot with very little powder that travels only a short distance, and successive shots with successively more powder until a circular orbit is obtained. More powder leads to more velocity, v, more specific angular momentum, h, more specific energy, $\mathcal{E}$, larger semi-latus rectum, p, larger semi-major axis, a, smaller eccentricity, e, and less semi-distance between the foci, c. The circle may be thought of as the limiting case between interior elliptical orbits and exterior ones. Compared with the interior orbits the circular orbit has the maximum values of $v, h, \mathcal{E}, a$, and p and the minimum values of e and c. For all of these interior elliptical orbits the cannonball starts at apoapsis (or apogee).

Loading even more powder would lead to successively larger exterior orbits. The cannonball would start at periapsis (or perigee). The circular orbit would represent a limiting case to the exterior orbits and would have the minimum values of $v, h, \mathcal{E}, a, p, e$, and c. Using even more powder would eventually result in a parabolic orbit and a family of hyperbolic orbits. The circular orbit would remain the minimum case for this entire family of orbits.

Degenerate Conics

Given the definition of a conic section as the intersection of a plane with a cone, I feel obliged to point out that such intersections include a single point at the apex, which is also the force center, a single line that includes the force center, and pairs of lines that intersect at the force center. From a practical sense these orbits have little value to space travel because of their intersection with the force center. The point may be interpreted as a static object at the force center. The line orbits could be interesting if we could have a straight-line tunnel through the Earth that includes the Earth's center. The straight-line orbit could also be terrifying if we identify an object that is on such an orbit and we live on the force center.

Problem 7 A tunnel through the Earth

Assume the Earth to be spherical. Imagine a tunnel through the Earth's diameter, which, by definition, includes the Earth's center. Hypothetically, it would be possible to jump into such a tunnel and get free transportation to the other side of the Earth, courtesy of gravity. What would be the values of $\mathcal{E}, \vec{h}, p, e$, and a for such an orbit? What kind of conic would it be? How long would it take to get to the other side of the Earth?

To complete this problem it helps to think of the Earth as made of concentric spherical shells. The shells outside of your current position in the tunnel exert a net gravitational force of zero - only the interior shells matter. Neglect air resistance in the tunnel and assume that the Earth has uniform denisty.

1.6.4 What is $\vec{B}$?

The vector constant of integration is at best unfamiliar. It is obviously important because it provides the reference direction from which the angle

ν is measured, as shown in Figure 1.5-1 in the text. Our derivation of the conic sections gives us a hint of the direction of $\vec{B}$, for in those derivations the angle θ was measured from the horizontal. Going back a few steps, we can solve the vector form of the trajectory equation to get

$$\vec{B} = \dot{\vec{r}} \times \vec{h} - \mu \frac{\vec{r}}{r}. \tag{1.139}$$

My first inclination is to determine $\vec{B}$ for the simplest case, which seems to be a circle. In this case $\vec{r}$ and $\vec{v}$ are always perpendicular, so the magnitude of $\vec{h}$ is rv and the cross product $\dot{\vec{r}} \times \vec{h}$ has magnitude rv^2 and its direction is always outward from the center of the circular orbit. The direction of $-\mu\frac{\vec{r}}{r}$ is always inward, so we only have to determine its magnitude relative to rv^2. The vector $\frac{\vec{r}}{r}$ is a unit vector, so we simply need to know the magnitude of μ, which is GM. We already know that for the circular orbit $v_{cs} = \sqrt{\frac{\mu}{r_{cs}}}$, so

$$r_{cs} v_{cs}^2 = r_{cs} \frac{\mu}{r_{cs}} = \mu, \tag{1.140}$$

which gives $\vec{B} = 0$! We have chosen the special case in which $\vec{B}$ is zero. In retrospect, this makes sense, because a circle has no obvious reference direction.

Next consider an elliptical orbit external to the circle, and specifically examine the orbit at pericenter. Here $\vec{r}$ and $\vec{v}$ are still perpendicular, and $\dot{\vec{r}}$ is still perpendicular to $\vec{h}$, but $v > v_{cs}$ at that radius, so $\vec{B}$ points outward, away from the focus. Thus, $\vec{B}$ points in the direction from the focus to pericenter.

Finally, consider an elliptical orbit internal to the circle. In this case the vector reverses because $v < v_{cs}$, but the far focus is occupied, so the net again points toward pericenter.

Problem 8 More on the direction of $\vec{B}$

Convince yourself that $\vec{B}$ is toward pericenter for parabolic and hyperbolic orbits.

In many advanced mechanics books $\vec{B}$ is shown to be closely related to the Laplace-Runge-Lenz vector, specifically, $\vec{B}$ is that vector divided by the mass, so we might give $\vec{B}$ the clumsy name specific Laplace-Runge-Lenz vector. Clearly, $\vec{B}$ is an additional constant of the motion. $\vec{B}$ depends on $\dot{\vec{r}}$, as does $\mathcal{E}$, and it depends on $\vec{h}$, so it does not add three vector components that are all constants of the motion, only one. For more on this vector see page 103 of *Classical Mechanics*, third edition, by Goldstein, Poole, and Safko, published by Addison Wesley, copyright 2002.

1.6.5 The Eccentricity Vector

Following the above, we define

$$\vec{e} = \vec{B}/\mu, \tag{1.141}$$

the eccentricity vector, which points toward pericenter and has magnitude equal to the eccentricity. This vector will simplify calculation of the orbital elements in Chapter 2 and beyond. Working from the definition of $\vec{B}$ we find

$$\mu\vec{e} = \vec{B} = \vec{v} \times \vec{h} - \mu\frac{\vec{r}}{r} = \vec{v} \times (\vec{r} \times \vec{v}) - \mu\frac{\vec{r}}{r} = v^2\vec{r} - (\vec{v} \cdot \vec{r})\vec{v} - \mu\frac{\vec{r}}{r}, \tag{1.142}$$

where the last step makes use of the $BAC - CAB$ rule. This result will be useful in future calculations.

1.7 Canonical Units

The text has been superseded since it was published in 1971. Now, astronomical distances and masses *are* well known. Nonetheless, canonical units remain in use for convenience, and because every discipline uses units appropriate to the problems that it investigates. In many cases these are not SI units, and canonical units are not.

Please refer back to the problem titled "The Mass of the Sun." In the years after World War II, when captured German rocket technology made it clear that space travel would occur, the mean radius of the Earth's orbit - the Astronomical Unit - was not known with acceptable certainty, so the mass of the Sun was not certain enough for detailed space-mission planning. Planning was done in canonical units, and this continues because of their simplicity.

The mathematical form of Kepler's Third Law says

$$T^2 = \frac{4\pi^2}{\mu}a^3. \tag{1.143}$$

If G is certain and M is uncertain, but Solar orbits with great relative accuracy must be calculated for mission planning, then choose a system of units in which $\mu = GM$ can be well known. Choose the Solar mass to be exactly 1 mass unit, and choose the length of measuring sticks and the rate of clocks so that G can also be exactly 1. Then, whenever the Solar mass is accurately known all masses may be scaled to the accurate value, and the measuring sticks and clock rates can be similarly re-scaled. We note that having $\mu = 1$ will greatly simplify the calculations of $\mathcal{E}$ from the energy equation, r from the equation of motion, and many other quantities describing orbits, such as p and e. Solving the above equation for μ we find

$$\mu = 4\pi^2 \frac{a^3}{T^2}. \tag{1.144}$$

The leading term of $4\pi^2$ is a dimensionless constant, so in any system of units μ must have the dimensions of length cubed over time squared, matching Kepler's Third Law.

Choose the distance unit, DU, and the time unit, TU, so that $\mu = 1\frac{DU^3}{TU^2}$. To do this, first choose a circular reference orbit. For heliocentric orbits this is a circular orbit with a radius of one astronomical unit, or $1AU$, so $DU = 1AU$. Now choose a time unit that makes everything tie together as desired. To do this, choose a time unit that gives the object in the reference orbit unit speed, so $DU/TU = 1$. We already have the ability to calculate the orbital speed as

$$v = \frac{2\pi DU}{T}, \tag{1.145}$$

where $2\pi DU$ is the circumference of the circular orbit and T is its period. The two velocity definitions become consistent if

$$TU = \frac{T}{2\pi}. \tag{1.146}$$

Let's derive the quantities in the text. The period of the Earth's orbit about the Sun, T, is 365.24 days, which works out to be 3.15567×10^7 seconds, so $TU = T/2\pi = 5.02241 \times 10^6$ seconds, which agrees with the values in Appendix A of the text to four significant figures. The Astronomical Unit is 1.4959965×10^8 kilometers, so the speed unit, $DU/TU = 29.7862$ kilometers per second, which also agrees acceptably with the text. Following scientific style we will write kilometers per second as km s^{-1}, and proceed similarly.

The gravitational parameter is $\mu = DU^3/TU^2 = 1.32729 \times 10^{11}$ km^3 s^{-2}, still consistent with the text to four significant figures.

For any other gravitating body the reference orbit is a circular orbit that just grazes its mean surface. The distance unit is the radius of this circular orbit. The time unit is still chosen so that the speed of the object in this orbit is $1DU/TU$, and μ is $1DU^3/TU^2$. For Earth orbit Appendix A of our text shows that $1DU = 2.09256 \times 10^7$ feet = 3963.19 miles = 3443.92 nautical miles = 6378.14 kilometers. Again we choose the time unit so that $TU = T/2\pi$, but what is T?

By the year 1800 the circumference of the Earth was reasonably well known from a survey done by J. B. J. Delambre and P. F. A. Mechain, and the density of the Earth was well known from the work of Henry Cavendish. Cavendish's work also enabled the calculation of the gravitational constant, G, although this was not done until 1873 by A. Cornu and J. B. Baille. By the time the Earth data were recognized as needed they were already available, while the solar data still were not. From our study of the circular orbit as a conic section we know

$$T = \frac{2\pi}{\sqrt{\mu}} r^{3/2}, \tag{1.147}$$

but $TU = T/2\pi$, so

$$TU = \frac{T}{2\pi} = \frac{r^{3/2}}{\sqrt{\mu}}. \tag{1.148}$$

Thus, μ and TU can be calculated from first principles using $G = 6.67 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, $M = 5.98 \times 10^{24}$ kg, and $r = DU = 6,378,000$ m, to get $\mu = 3.98866 \times 10^{14}$ m^3 s^{-2} or 3.98866×10^5 km^3 s^{-2} and $TU = 805.171$ s. The values of 3.986012×10^5 km^3 s^{-2} and 806.811 s in the text make use of newer values of G, M, and DU, but the results agree very closely. The associated speed unit is 7.9215 km s^{-1}, which also agrees well with the tabulated value of 7.90536 km s^{-1}.

It would be pleasing if things could stay so simple, but they cannot, for the metric system devised by the French at the time of their revolution was not well accepted in the English-speaking world. The meter was defined as 10^{-7} times the distance from the equator to the pole at sea level. The English, and by historical association Americans, used the nautical mile, which is one arc minute of latitude. Consequently, we still use feet and nautical miles, somewhat to our peril.

Please see

www.sizes.com/units/meter.htm

www.sizes.com/units/mile_nautical.htm and

www.cnn.com/TECH/space/9909/30/mars.metric.02
for more on these topics.

As an example, let's work the problem on page 41 of the text.

Example 1 A Cooked Up Example

A space object is observed at an altitude of 1.046284×10^7 feet above the Earth traveling at 2.593625×10^4 feet s^{-1} with a flight-path angle of 0^o. Using canonical units determine $\mathcal{E}, h, p, e, r_a$, and r_p.

Appendix A shows that in Earth units $1DU = 2.092567257 \times 10^7$ feet and $1DU/TU = 25936.24764$ feet s^{-1}. This means that the object is $0.5000000DU$ above the surface of the Earth and is traveling with a speed of $1.000000DU/TU$ on a flight-path angle of zero. The orbital radius is then the radius of the Earth plus 0.5000000 times the radius of the Earth, or exactly $1.5DU$ in canonical units. We calculate that

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} = \frac{1}{2} - \frac{1}{1.5} = \frac{1}{2} - \frac{2}{3} = -\frac{1}{6} = -0.16666DU^2/TU^2. \tag{1.149}$$

Similarly,

$$h = rv\cos\phi = 1.5DU^2/TU. \tag{1.150}$$

The parameter is

$$p = \frac{h^2}{\mu} = \frac{2.25DU^4/TU^2}{1DU^3/TU^2} = 2.25DU. \tag{1.151}$$

The eccentricity is

$$e = \sqrt{1 + \frac{2\mathcal{E}h^2}{\mu^2}} = \sqrt{1 + \frac{-0.3333 \times 2.25}{1}} = \sqrt{1 - \frac{3}{4}} = \sqrt{\frac{1}{4}} = \frac{1}{2}. \tag{1.152}$$

The radius at apogee is

$$r_a = \frac{p}{1-e} = \frac{2.25}{0.5} = 4.5DU. \tag{1.153}$$

The radius at perigee is

$$r_p = \frac{p}{1+e} = \frac{2.25}{1.5} = 1.5DU. \tag{1.154}$$

I choose to give the authors a lot of credit for a clear example, but otherwise this has some problems with error bars and significant figures.

Chapter 2

Orbit Determination from Observations

The existence of Kepler's Laws indicates to us that Kepler determined orbits, and Newton published a method in the Principia. This chapter develops the modern approach to orbit determination.

2.1 Coordinate Systems

We seek an inertial coordinate system in which to apply the results of Chapter 1 before we can proceed. To specify a coordinate system we must give the location of the origin, the orientation of the fundamental plane (the $X-Y$ plane), the principal direction within the fundamental plane (the direction of X), and the direction of the Z axis. Right-handed coordinates are assumed. There may be more than one such coordinate system, so we will also develop the mathematics for transforming from one system to another.

In reality none of the coordinate systems that we will define are truly inertial, because all are accelerating to some degree. In many cases the accelerations are small enough that the coordinates are good enough for practical purposes.

We will start with four coordinate systems and add more when we need them. The four systems are the Heliocentric-Ecliptic coordinate system, the Geocentric-Equatorial coordinate system, the Right Ascension-Declination coordinate system, and the Perifocal coordinate system. With experience it becomes clear that the names at least hint at the location of the origin and the orientation of the fundamental plane for each system.

The Celestial Sphere is important to most celestial coordinate systems. This is a sphere of essentially infinite diameter onto which the stars, which are taken to be at infinite distance, are projected. Thus, the apparent locations of the stars are taken to be independent of the location of the observer on the Earth. Objects of the Solar System and orbiting bodies are also seen projected onto the Celestial Sphere, but because these objects are close to the Earth their apparent locations on the Celestial Sphere depend on the location of the observer.

2.1.1 Heliocentric-Ecliptic Coordinates

As the name implies, Heliocentric-Ecliptic coordinates have their origin at the center of the Sun. They are most useful for describing orbits around the Sun. Their fundamental plane is the plane of the ecliptic, which is the plane of the Earth's orbit. This is an infinite plane, extending to and intersecting the Celestial Sphere. Within the plane of the ecliptic we can begin to define the fundamental directions by determining the line of intersection between the ecliptic plane and the Earth's equatorial plane. A line from the center of the Earth to the center of the Sun at the start of Northern Hemisphere spring defines the fundamental direction along this line, and its intersection with the Celestial Sphere is the First Point of Aries, also called the Vernal Equinox. This direction is labeled X_ϵ. The Z_ϵ direction is along the direction of the Earth's orbital angular momentum vector. This puts the Y_ϵ direction toward the line from the center of the Sun to the center of the Earth at the start of winter.

The mass distribution of the Earth is not perfectly uniform, so these directions precess slowly from torques applied primarily by the Sun and Moon, and their actual directions must be specified for a given year, called an epoch, i.e., epoch 2000. Today the First Point of Aries is actually in the constellation Pisces. Now you know why an astronomer is teaching orbital mechanics to engineers.

A closely related coordinate system is the Solar System Barycentric coordinate system. This uses the same fundamental directions, but its origin is at the center of mass of the Solar System, which is near the surface of the Sun in the direction of Jupiter.

2.1.2 Geocentric-Equatorial Coordinates

These coordinates have their origin at the center of the Earth. They are used to describe orbits around the Earth. The fundamental plane is that of the Earth's equator, which is extended to intersect the Celestial Sphere at the Celestial Equator. These coordinates *do not* rotate with the Earth. This will force us to be keenly aware of our location on the Earth and the time, for Geocentric-Equatorial coordinates depend on them. The X direction is toward the Vernal Equinox, the same as X_ϵ. The Z direction points toward the North Celectial Pole, near the North Star. There is an associated South Celestial Pole for the benefit of those in the Southern Hemisphere. I have seen it an it is real. The Y direction is perpendicular to the X and Z directions. These directions are specified by unit vectors $\hat{I}$, $\hat{J}$, and $\hat{K}$, respectively.

The Y and Z directions are not the same as those in the Geocentric-Equatorial coordinates because the rotational angular momentum vector of the Earth is inclined with respect to the orbital angular momentum vector.

2.1.3 Right Ascension-Declination Coordinates

These are closely related to Geocentric-Equatorial coordinates and use the same fundamental directions. They are used to describe the coordinates of stars and galaxies, which appear relatively fixed except for precession, and for planets, which move because of their nearby location and orbital motion. Satellites, asteroids, and comets also appear to move in this system for the same reasons of proximity and orbital motion. Their positions are often determined by their locations relative to nearby stars, proving the utility and importance of this system.

Right Ascension and Declination are both specified by angles - Right Ascension in the plane of the Celestial Equator eastward from $\hat{I}$ and Declination northward from the Celestial Equator.

The origin of these coordinates may be at the center of the Earth or somewhere on the surface of the Earth, called the topos. For nearby objects the coordinates depend on the location of the origin. These coordinates do not rotate with the Earth.

2.1.4 Perifocal Coordinates

Perifocal coordinates can apply to any two-body gravitating system. The origin of Perifocal coordinates is the focus occupied by the central gravitating body. The fundamental plane is the plane of the orbiting body, and the fundamental direction is the direction from the gravitating body to pericenter. It is *very* easy to describe an orbit in this system. The coordinate from the force center to pericenter is labeled x_ω. The coordinate ninety degrees away, in the plane of the orbit and in the direction of the motion of the orbiting body, is labeled y_ω. The coordinate in the direction of the angular momentum vector, $\vec{h}$, is labeled z_ω. The respective unit vectors are called $\hat{P}, \hat{Q}$, and $\hat{W}$. $\hat{P}$ is a unit vector in the direction of $\vec{e}$ and $\hat{W}$ is a unit vecrot in the direction of $\vec{h}$.

2.2 Classical Orbital Elements

Five independent quantities, called orbital elements are sufficient to describe the size, shape, and orientation of an orbit. One more is needed to locate the orbiting body at a particular place and time. The classical orbital elements are:

a, the semi-major axis, a constant describing the size of the orbit,

e, the eccentricity, a constant defining the shape of the orbit,

i, the inclination, the angle between $\vec{h}$ and $\hat{K}$,

Ω, the longitude of the ascending node, an angle in the fundamental plane between $\hat{I}$ and the ascending node of the line of nodes, measured counterclockwise when viewed from the north side of the fundamental plane,

ω, the argument of periapsis, an angle in the orbital plane between the ascending node and periapsis, measured in the direction of the satellite's motion, and

T, the time of periapsis passage.

Before going further, this is a good time to explain the naming of angles. If the measure of an angle starts from $\hat{I}$ then the angle is called a longitude. If the measure of an angle starts from $\vec{n}$ then the angle is called an argument.

There are important cases in which some of the quantities are not defined. For instance, in a circular orbit there is no periapsis, so ω is not defined. In an orbit of zero inclination the fundamental and orbital planes coincide, so there is no ascending node and Ω is undefined. This coincidence is also called an equatorial orbit.

This list is sufficient, but not exhaustive. We saw in Chapter 1 that there is a relation among a, e, and p, so any two can be used to determine the third. Thus, p is often substituted for a. Modern radar observations make it easier to find p than a for most Solar System objects.

There is also a good substitute for ω, the argument of periapsis, which is Π, the longitude of periapsis. This is the angle from $\hat{I}$ measured in the fundamental plane eastward to the ascending node plus the angle from there to periapsis in the orbital plane. This is a very unusual physical quantity because it is the sum of two angles in different planes. If both Ω and ω are defined then $\Pi = \Omega + \omega$. If ω is not defined then all is not lost, it still may be possible to define Π. As long as there is a periapsis, meaning that the orbit is not circular, Π is the angle between $\hat{I}$ and $\vec{e}$.

It is also possible to substitute for the time of periapsis passage, usually by specifying the location of the satellite at the time of observation, t_o, also called the epoch. The true anomaly at epoch, ν_o, is the angle in the orbital plane between periapsis and the position of the satellite at t_o. This is the angle that we called ν in deriving the trajectory equation, so it may be familiar. The subscript is meant to indicate that the angle has been observed at the particular time t_o. The argument of latitude at epoch, u_o, is the angle in the orbital plane between the ascending node and the location of the satellite at t_o. For nonequatorial orbits $u_o = \omega + \nu_o$. For equatorial orbits u_o is undefined. The true longitude at epoch, ℓ_o, is the angle from $\hat{I}$ measured eastward in the fundamental plane to the ascending node plus the angle measured in the orbital plane, in the direction of motion, to the location of the satellite. If Ω, ω, and ν_o are all defined then

$$\ell_o = \Omega + \omega + \nu_o = \Pi + \nu_o = \Omega + u_o. \tag{2.1}$$

There is no ascending node in an equatorial orbit, so Ω is undefined, and $\ell_o = \Pi + \nu_o$. There is no periapsis in a circular orbit, so ω is undefined, and $\ell_o = \Omega + u_o$. If the orbit is equatorial and circular then ℓ_o is the angle from $\hat{I}$ to the position of the satellite, which is always defined. We note that ℓ_o is an unusual physical quantity in the same way that Π is, for both are the sums of angles in two different planes.

Ω, Π, and ℓ_o are measured starting at $\hat{I}$, so theya re called longitudes. ω and u_o are measured starting at $\vec{n}$, so theya re called arguments. The true anomaly at epoch is measured starting at $\vec{e}$, so it is neither a longitude nor an argument.

Orbits are commonly called direct or retrograde. Direct orbits move to the east and retrograde orbits move to the west. I note that the Earth orbits to the east and spins to the east, as do most of the planets. Another way to say this is that most of the angular momentum vectors in the Solar System point north, or nearly so. It takes less energy to launch an artificial satellite into a direct orbit than a retrograde one for the same reason - direct orbits take advantage of the existing motion, retrograde orbits must first overcome it and then establish the desired motion in the opposite direction.

2.3 Determining the Orbital Elements from $\vec{r}$ and $\vec{v}$

Radar sites can provide the position and velocity vectors, $\vec{r}$ and $\vec{v}$, at the epoch of observation, t_o, for satellites and most of the planets. Let us start there.

2.3.1 Three Fundamental Vectors

The orbital elements are calculated by first finding the angular momentum and eccentricity vectors, $\vec{h}$ and $\vec{e}$, defined in Chapter 1, plus the node vector, $\vec{n}$, which is new to us.

By now the specific angular momentum should be familiar,

$$\vec{h} = \vec{r} \times \vec{v}. \tag{2.2}$$

Its length is the specific angular momentum and its direction is perpendicular to the orbit.

The eccentricity vector is

$$\vec{e} = \frac{1}{\mu}\left[\left(v^2 - \frac{\mu}{r}\right)\vec{r} - (\vec{r} \cdot \vec{v})\vec{v}\right]. \tag{2.3}$$

Its length is the eccentricity and its direction is from the occupied focus toward periapsis.

The node vector is

$$\vec{n} = \vec{K} \times \vec{h}. \tag{2.4}$$

The length of $\vec{n}$ contains no new information and is not of interest. From the definition of the cross product $\vec{n}$ must be perpendicular to both $\vec{K}$ and $\vec{h}$. Being perpendicular to $\vec{K}$ places $\vec{n}$ in the fundamental (or equatorial) plane. Being perpendicular to $\vec{h}$ places $\vec{n}$ in the orbital plane. Thus, it must be at the intersection of those planes, which is the line of nodes. The order of the cross product is chosen so that $\vec{n}$ points from the force center to the ascending node.

2.3.2 Solving for the Orbital Elements

While solving for the orbital elements it is important to become familiar with Figure 2.3-1 in the text. There are seven important vectors that are drawn in the figure and one that is not drawn. The three basis vectors, $\hat{I}, \hat{J}$, and $\hat{K}$, are the unit vectors of the coordinate system, in this case the Geocentric-Equatorial system. Their origin is the center of the Earth. $\hat{I}$ points toward the vernal equinox, $\hat{K}$ points toward the North Celestial Pole, and $\hat{J}$ points toward the first point of winter. $\hat{I}$ and $\hat{J}$ are in the fundamental plane, in this case the equatorial plane, and $\hat{K}$ is perpendicular to it. These coordinates do not rotate with the Earth, but their origin does remain at the Earth's center as the Earth orbits the Sun. The three vectors $\vec{h}, \vec{e}$, and $\vec{n}$ describe the orbit. They share the same origin with the basis vectors. The specific angular momentum vector, $\vec{h}$, is perpendicular to the orbital plane, the eccentricity vector, $\vec{e}$, points toward perigee, and the node vector, $\vec{n}$ points toward the ascending node. The vector $\vec{r}$ describes the current position of the satellite. The vector that is not drawn is $\vec{v}$. It is tangent to the orbit and describes the current velocity. The five vectors $\vec{h}, \vec{e}, \vec{n}, \vec{r}$, and $\vec{v}$ share the origin with the basis vectors. In a true two-body problem $\vec{h}, \vec{e}$, and $\vec{n}$ do not change with time, although $\vec{r}$ and $\vec{v}$ do.

We already know how to find p from $\vec{h}$, and $\vec{e}$ easily gives us its own magnitude, e. If a is desired it can be calculated from p and e. The remaining quantities are angles between pairs of vectors illustrated in the text. If we can find the angles given the vectors then we are home free. Remember, from your earliest knowledge of vectors, that given vectors $\vec{A}$ and $\vec{B}$ separated by an angle α,

$$\vec{A} \cdot \vec{B} = AB \cos\alpha, \tag{2.5}$$

so

$$\cos\alpha = \frac{\vec{A} \cdot \vec{B}}{AB}. \tag{2.6}$$

The particular orientations of the vectors will be used to determine whether the angle is less than or greater than 180^o. We now proceed to find the orbital elements, given $\vec{h}$, $\vec{e}$, and $\vec{n}$.

First, frrom $\vec{h}$,

$$p = \frac{h^2}{\mu}. \tag{2.7}$$

Next,

$$e = |\vec{e}|. \tag{2.8}$$

If a is needed it can be calculated from $a = p/(1 - e^2)$.

By definition, i is the angle between $\hat{K}$ and $\vec{h}$, so

$$\cos i = \frac{h_K}{h} \tag{2.9}$$

because $\hat{K}$ is a unit vector. The inclination is always less than 180^o.

Since Ω is the angle from $\hat{I}$ to $\vec{n}$ it is called the longitude of the ascending node, and

$$\cos \Omega = \frac{\hat{I} \cdot \vec{n}}{n} = \frac{n_I}{n}. \tag{2.10}$$

If $n_J > 0$ then Ω is less than 180^o.

Since ω is the angle from $\vec{n}$ to $\vec{e}$ it is called the argument of periapsis, and

$$\cos \omega = \frac{\vec{n} \cdot \vec{e}}{ne}. \tag{2.11}$$

If $e_K > 0$ then ω is less than 180^o.

Since ν_o is the angle between $\vec{e}$ and $\vec{r}$ at time t_o it is called the true anomaly at epocy, and

$$\cos \nu_o = \frac{\vec{e} \cdot \vec{r}}{er}. \tag{2.12}$$

If $\vec{r} \cdot \vec{v} > 0$ then ν_o is less than 180^o.

Since u_o is the angle between $\vec{n}$ and $\vec{r}$ at time t_o it is called the argument of latitude at epoch, and the

$$\cos u_o = \frac{\vec{n} \cdot \vec{r}}{nr}. \tag{2.13}$$

If $r_K > 0$ then u_o is less than 180^o.

Since ℓ_o is the sum of other angles starting at $\hat{I}$ and finishing at $\vec{r}$ at time t_o it is called the true longitude at epoch, and

$$\ell_o = \Omega + \omega + \nu_o = \Omega + u_o. \tag{2.14}$$

The angle-naming system could be made more consistent if u_o were called the true argument of latitude ate epoch. Then every angle beginning at $\hat{I}$ would be a longitude, every angle beginning at $\vec{n}$ would be an argument, and every angle ending at $\vec{r}$ would be true.

Please study the examples that follow in the text in Figures 2.4-2, 2.4-3, and 2.4-4. Note that the orbit is drawn from the orbital elements - you are not expected to figure out the orbital elements from the drawing.

Example 2 A Simple Case of Calculating Orbital Elements

A radar tracks a satellite and its computer system converts the position and velocity vectors to Canonical units, giving

$$\vec{r} = 2\hat{I}DU, \vec{v} = 1\hat{J}DU/TU. \tag{2.15}$$

Of course, if a computer system can be programmed do that it can also be programmed to calculate the orbital elements. Please put that aside in the interest of developing physical intuition, and find the orbital elements. First, note that $\vec{r}$ and $\vec{v}$ are perpendicular. This tells us that either the orbit is a circle or the satellite is at a special place in its orbit, such as perigee or apogee - physical intuition!

Starting as directed we first find $\vec{h}, \vec{e}$, and $\vec{n}$.

$$\vec{h} = \vec{r} \times \vec{v} = 2\hat{K}\frac{DU^2}{TU}, \tag{2.16}$$

$$\vec{e} = \frac{1}{\mu}\left[\left(v^2 - \frac{\mu}{r}\right)\vec{r} - (\vec{r}\cdot\vec{v})\vec{v}\right] = 1\frac{TU^2}{DU^3}\left[\left(1 - \frac{1}{2}\right)\frac{DU^2}{TU^2}2\hat{I}DU - (0)\right] = 1\hat{I}, \tag{2.17}$$

$$\vec{n} = \hat{K} \times \vec{h} = 0. \tag{2.18}$$

We have interesting results already. $\vec{h}$ is in the same direction as $\hat{K}$, so the orbit is equatorial, so there is no ascending node, and $\vec{n}$ is zero. We also see that $\vec{e}$ is dimensionless.

Proceeding,

$$p = \frac{h^2}{\mu} = 4\frac{DU^4}{TU^2}\frac{TU^2}{DU^3} = 4DU, \tag{2.19}$$

$$e = |\vec{e}| = 1. \tag{2.20}$$

Since $e = 1$ but $h \neq 0$ the orbit is a parabola and we cannot use $a = \frac{p}{1-e^2}$ to find a. As e approaches one a approaches infinity, which is correct for a parabola.

The inclination is the angle between $\vec{h}$ and $\hat{K}$, but they are parallel, so i should be zero. Let's confirm this.

$$i = \arccos\frac{\vec{h}\cdot\hat{K}}{h} = \arccos\frac{h}{h} = \arccos 1, \tag{2.21}$$

and since i is always less than 180^o, $i = 0$. This confirms that the orbital plane and the fundamental plane coincide, so the orbit is equatorial. Note that there is a misprint in the text, which uses $\vec{k}$ instead of $\hat{K}$.

The longitude of the ascending node is

$$\Omega = \arccos\frac{\vec{n}\cdot\hat{I}}{n}, \tag{2.22}$$

but $\vec{n} = 0$, so Ω is undefined. This confirms that there is no line of nodes.

The same holds true for ω, the argument of the periapsis; there is no line of nodes because $\vec{n} = 0$, so ω is undefined.

Next find the longitude of periapsis, Π. At the risk of repeating myself, the orbital plane is coincident with the equatorial plane, there is no line of nodes, so Π is measured from the $\hat{I}$ axis to periapsis, which is in the direction of $\vec{e}$.

$$\Pi = \arccos\frac{\vec{e}\cdot\hat{I}}{e} = \arccos\frac{\hat{I}\cdot\hat{I}}{1} = \arccos 1, \tag{2.23}$$

We take $\Pi = 0^o$. Perigee is on the $\hat{I}$ axis, and the satellite was at perigee at the time of the observation.

The true anomaly is

$$\nu_o = \arccos\frac{\vec{e}\cdot\vec{r}}{er} = \arccos 1. \tag{2.24}$$

The satellite was at perigee at the time of observation, so $\nu_0 = 0^o$.

Finally,

$$\ell_0 = \Pi + \nu_o = 0^o. \tag{2.25}$$

This example showed how easy these calculations can be if the radius and velocity are in canonical units with happy numbers and the satellite is at perigee. We were able to guess some of the properties of the orbital elements.

We should ask, will the satellite hit the Earth? The perigee is $2DU$ and the parameter is $p = 4DU > 1DU$, so it will not hit the Earth. This is another example of the utility of canonical units. We can now convert the value of p, and all the others, into any system of units that we like.

The next example, that begins on page 70 of the text, earns the same marks. How many times have you been marked off for too many significant figures or no formal calculation of the uncertainty?

2.4 Determining $\vec{r}$ and $\vec{v}$ from the Orbital Elements

This is the inverse of the previous section. It becomes especially important in more advanced orbital mechanics for improving the calculation of future values of $\vec{r}$ and $\vec{v}$, which would enable the calculation of improved values of the orbital elements.

We make use of the ease of describing an obit in the Perifocal system, then we transform to Geocentric-Equatorial coordinates. Remember that in Perifocal coordinates the unit vectors are $\hat{P}$ and $\hat{Q}$ in the plane of the orbit and $\hat{W}$ perpendicular to it. Given the five orbital elements p, e, i, Ω, and ω, we can find $\vec{r}$ from a simple projection,

$$\vec{r} = r\cos\nu\hat{P} + r\sin\nu\hat{Q}, \tag{2.26}$$

where r is calculated from the trajectory equation,

$$r = \frac{p}{1 + e\cos\nu}. \tag{2.27}$$

Then we can find $\vec{v}$ by taking a time derivative,

$$\vec{v} = \dot{\vec{r}} = (\dot{r}\cos\nu - r\dot{\nu}\sin\nu)\hat{P} + (\dot{r}\sin\nu + r\dot{\nu}\cos\nu)\hat{Q}. \tag{2.28}$$

Our goal is to simplify this expression. We already know $h = r^2\dot{\nu}$ and $p = h^2/\mu$, and we can find $\dot{r}$ from the trajectory equation,

$$\dot{r} = p(-1)(1 + e\cos\nu)^{-2}(-e\sin\nu\dot{\nu}) = \frac{pe\dot{\nu}\sin\nu}{(1 + e\cos\nu)^2}. \tag{2.29}$$

This can be simplified because squaring the trajectory equation gives

$$r^2 = \frac{p^2}{(1 + e\cos\nu)^2}, \tag{2.30}$$

so

$$\dot{r} = \frac{r^2 e\dot{\nu}\sin\nu}{p} = \frac{h}{p}e\sin\nu = \sqrt{\frac{\mu}{p}}e\sin\nu. \tag{2.31}$$

We also have

$$r\dot{\nu} = \frac{h}{r} = \frac{h}{p}(1 + e\cos\nu) = \frac{\sqrt{p\mu}}{p}(1 + e\cos\nu) = \sqrt{\frac{\mu}{p}}(1 + e\cos\nu). \tag{2.32}$$

Substituting for both $\dot{r}$ and $r\dot{\nu}$ allows us to simplify the expression for $\vec{v}$ to

$$\vec{v} = \sqrt{\frac{\mu}{p}}\left[-\sin\nu\hat{P} + (e + \cos\nu)\hat{Q}\right], \tag{2.33}$$

which is the desired result.

Example 3 Another Happy Example

A space tracking station provides the following orbital elements: $p = 2.25DU, e = 0.5, i = 45^o, \Omega = 30^o, \omega = 0^o$, and $\nu_o = 0^o$. Use them to determine $\vec{r}$ and $\vec{v}$.

To apply the equations above we need the magnitude of $\vec{r}$, which can be found from the orbital elements and the trajectory equation,

$$r = \frac{p}{1 + e\cos\nu} = \frac{2.25DU}{1.5} = 1.5DU. \tag{2.34}$$

Then

$$\vec{r} = r\cos\nu\hat{P} + r\sin\nu\hat{Q} = 1.5\hat{P}DU, \tag{2.35}$$

and

$$\vec{v} = \sqrt{\frac{\mu}{p}}\left[-\sin\nu\hat{P} + (e + \cos\nu)\hat{Q}\right] = \sqrt{\frac{1}{2.25}}\left[0\hat{P} + 1.5\hat{Q}\right] = 1\hat{Q}DU/TU. \tag{2.36}$$

2.5 Coordinate Transformations

We wish to transform from Pericentric coordinates to Geocentric-Equatorial coordinates. We are unlikely to have lasting success at this if we do not understand what coordinate transformations are and how they work, so we will spend some time doing so.

Specifically, we wish to transform the position and velocity vectors, $\vec{r}$ and $\vec{v}$. We can and should think of these vectors as being fixed for any particular instant of time. Transforming them from one coordinate system to another does not change them, it merely changes the framework used in their description.

For each coordinate system we will specify an origin and three mutually perpendicular, right-handed basis vectors of unit length. There are coordinate systems in which the basis vectors are not mutually perpendicular, but we choose to eliminate them from consideration in this course.

2.5.1 Transformations Change the Basis Vectors

Coordinate transformations change the basis vectors used to describe the vectors under study. The vectors remain the same, only the basis-vector framework for their description changes. We will start in the Perifocal coordinate system, whose origin is at the center of the gravitating body, in this case the Earth. Its basis vectors, $\hat{P}, \hat{Q}$, and $\hat{W}$, are chosen so that $\hat{P}$ is in the orbital plane and points toward perigee, $\hat{Q}$ is in the orbital plane and is 90^o from $\hat{P}$ in the direction of the orbital motion, and $\hat{W}$ is in the same direction as the specific angular momentum, $\vec{h}$. The basis vectors do not rotate with the Earth.

Let's think about a vector in this system - for simplicity of thought consider a position vector $\vec{r}$. How do we describe it in this system? We use vector dot products, and determine the dot product of $\vec{r}$ with each basis vector. The component along $\hat{P}$ is $r_P = \vec{r} \cdot \hat{P}$, that along $\hat{Q}$ is $r_Q = \vec{r} \cdot \hat{Q}$, and that along $\hat{W}$ is $r_W = \vec{r} \cdot \hat{W}$. The formal expression for the vector form of $\vec{r}$ is

$$\vec{r} = r_P \hat{P} + r_Q \hat{Q} + r_W \hat{W}. \tag{2.37}$$

We will transform to the Geocentric-Equatorial (G-E) coordinate system. Its origin is also at the center of the Earth, so the two coordinate systems share the same origin. This is helpful. Both are right handed, and both do not rotate with the Earth, which is helpful, too. All that we will have to do is rotate the coordinate systems to find the relation between them.

The G-E basis vectors, $\hat{I}$, $\hat{J}$, and $\hat{K}$, are chosen so that $\hat{I}$ points toward the Vernal Equinox, $\hat{J}$ points toward the first point of winter, and $\hat{K}$ points toward the North Celestial Pole.

To express the position vector $\vec{r}$ in this system we follow the same procedure as we did in Perifocal coordinates. The projections of $\vec{r}$ are $r_I = \vec{r}\cdot\hat{I}, r_J = \vec{r}\cdot\hat{J}$, and $r_K = \vec{r}\cdot\hat{K}$. Then the formal expression for $\vec{r}$ in the G-E coordinates is

$$\vec{r} = r_I\hat{I} + r_J\hat{J} + r_K\hat{K}. \tag{2.38}$$

How do we convert from (r_p, r_q, r_W) coordinates to (r_I, r_J, r_K) coordinates? How do we make the coordinate transformation happen?

2.5.2 Simple Transformations

Example 4 A Simple, Single Rotation

Let's start with the simplest possible case, which is rotation about a single axis. Imagine that we have the familiar Cartesian unit, basis vectors $\hat{x}, \hat{y}$, and $\hat{z}$. We rotate the coordinates about the $\hat{z}$ axis through a counterclockwise angle γ. Counterclockwise rotations, as viewed looking down the rotation axis, are taken to be positive, and clockwise negative. We imagine that copies of the original basis vectors $\hat{x}, \hat{y}, \hat{z}$ remain in place, and we also have the new rotated basis vectors $\hat{x}', \hat{y}', \hat{z}'$. The rotation about the $\hat{z}$ axis means that $\hat{z} = \hat{z}'$, so for any point the z coordinate does not change, and $z = z'$. The challenge is to figure out what happens in the $\hat{x} - \hat{y}$ plane. Specifically, we want to know how to calculate $\hat{x}', \hat{y}'$, and $\hat{z}'$ in terms of the original basis vectors $\hat{x}, \hat{y}$, and $\hat{z}$. We assert that any vector under study is the same vector in both systems, only the framework used to describe it changes. Thus, for any vector, $\vec{r}$,

$$\vec{r} = r_x\hat{x} + r_y\hat{y} + r_z\hat{z} = r_{x'}\hat{x}' + r_{y'}\hat{y}' + r_{z'}\hat{z}', \tag{2.39}$$

where

$$r_x = \vec{r}\cdot\hat{x}, r_y = \vec{r}\cdot\hat{y}, r_z = \vec{r}\cdot\hat{z}, \tag{2.40}$$

and

$$r_{x'} = \vec{r}\cdot\hat{x}', r_{y'} = \vec{r}\cdot\hat{y}', r_{z'} = \vec{r}\cdot\hat{z}'. \tag{2.41}$$

Let's follow this procedure for the unit vectors themselves. We project the new primed unit vectors onto the old unprimed ones and find

$$\hat{x}' = (\hat{x}' \cdot \hat{x})\hat{x} + (\hat{x}' \cdot \hat{y})\hat{y} + (\hat{x}' \cdot \hat{z})\hat{z} = \cos\gamma\hat{x} + \cos(90^o - \gamma)\hat{y} + \cos 90^o \hat{z}$$
$$= \cos\gamma\hat{x} + \sin\gamma\hat{y} + 0\hat{z}, \quad (2.42)$$

and

$$\hat{y}' = (\hat{y}' \cdot \hat{x})\hat{x} + (\hat{y}' \cdot \hat{y})\hat{y} + (\hat{y}' \cdot \hat{z})\hat{z} = \cos(90^o + \gamma)\hat{x} + \cos\gamma\hat{y} + \cos 90^o \hat{z}$$
$$= -\sin\gamma\hat{x} + \cos\gamma\hat{y} + 0\hat{z}, \quad (2.43)$$

and

$$\hat{z}' = (\hat{z}' \cdot \hat{x})\hat{x} + (\hat{z}' \cdot \hat{y})\hat{y} + (\hat{z}' \cdot \hat{z})\hat{z} = \cos 90^o\hat{x} + \cos 90^o\hat{y} + \cos 0^o \hat{z}$$
$$= 0\hat{x} + 0\hat{y} + 1\hat{z}. \quad (2.44)$$

Now plug these back into equation 2.39 to get

$$r_x\hat{x} + r_y\hat{y} + r_z\hat{z} = r_{x'}(\cos\gamma\hat{x} + \sin\gamma\hat{y}) + r_{y'}(-\sin\gamma\hat{x} + \cos\gamma\hat{y}) + r_{z'}\hat{z}$$
$$= (r_{x'}\cos\gamma - r_{y'}\sin\gamma)\hat{x} + (r_{x'}\sin\gamma + r_{y'}\cos\gamma)\hat{y} + r_{z'}\hat{z}. \quad (2.45)$$

Since the unit vectors are orthogonal we can equate just the parts specific to one unit vector to get

$$r_x = \cos\gamma r_{x'} - \sin\gamma r_{y'}, \quad (2.46)$$

$$r_y = \sin\gamma r_{x'} + \cos\gamma r_{y'}, \quad (2.47)$$

and

$$r_z = r_{z'}. \quad (2.48)$$

In matrix form this becomes

$$\begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{x'} \\ r_{y'} \\ r_{z'} \end{pmatrix}, \quad (2.49)$$

where we have written the transformation matrix.

This isn't quite what we want, which is the transformation going in the other direction - given the coordinates in the unprimed frame we want to calculate those in the primed frame. None the less, we have discovered how to proceed. We want the other projection of the vectors. Before we do this in detail let's look at what we have already done in the most general way.

Our transformation matrix can be written

$$\begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \hat{x}'\cdot\hat{x} & \hat{y}'\cdot\hat{x} & \hat{z}'\cdot\hat{x} \\ \hat{x}'\cdot\hat{y} & \hat{y}'\cdot\hat{y} & \hat{z}'\cdot\hat{y} \\ \hat{x}'\cdot\hat{z} & \hat{y}'\cdot\hat{z} & \hat{z}'\cdot\hat{z} \end{pmatrix}. \tag{2.50}$$

Temporarily call this matrix $\tilde{\mathbf{M}}$. We recognize that the desired projection is that of the unprimed unit vectors onto the primed ones. Let's do that, then exchange the order of the dot product, which does not change the result.

$$\begin{pmatrix} \hat{x}\cdot\hat{x}' & \hat{y}\cdot\hat{x}' & \hat{z}\cdot\hat{x}' \\ \hat{x}\cdot\hat{y}' & \hat{y}\cdot\hat{y}' & \hat{z}\cdot\hat{y}' \\ \hat{x}\cdot\hat{z}' & \hat{y}\cdot\hat{z}' & \hat{z}\cdot\hat{z}' \end{pmatrix} = \begin{pmatrix} \hat{x}'\cdot\hat{x} & \hat{x}'\cdot\hat{y} & \hat{x}'\cdot\hat{z} \\ \hat{y}'\cdot\hat{x} & \hat{y}'\cdot\hat{y} & \hat{y}'\cdot\hat{z} \\ \hat{z}'\cdot\hat{x} & \hat{z}'\cdot\hat{y} & \hat{z}'\cdot\hat{z} \end{pmatrix} = \tilde{\mathbf{M}}^{\mathbf{T}}. \tag{2.51}$$

We see that the new result is just the transpose of the previous one, and that it describes the transformation from the unprimed basis to the primed one, so

$$\begin{pmatrix} r_{x'} \\ r_{y'} \\ r_{z'} \end{pmatrix} = \tilde{\mathbf{M}}^{\mathbf{T}} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}. \tag{2.52}$$

We can now identify the matrix $\tilde{\mathbf{M}}^{\mathbf{T}}$ with the matrix $\tilde{\mathbf{C}}$ on page 78 of the text, so $\tilde{\mathbf{M}}$ must be its transpose, $\tilde{\mathbf{C}}^{\mathbf{T}}$.

The authors of our text do not make it clear why they use the tilde notation, as in $\tilde{\mathbf{C}}$. In some texts the tilde is used to indicate the transpose. That is not the case in our text, where the superscript T is used. I think that the tilde is used to indicate that the object is a matrix, or even a tensor, and that is how the tilde is used in these course notes.

Problem 9 Transpose, Inverse, and Orthogonal Matrices

If $\tilde{\mathbf{C}}$ performs the transformation from the unprimed coordinates to the primed ones, and $\tilde{\mathbf{C}}^{\mathbf{T}}$ performs the transformation from the primed coordinates to the unprimed ones, then we expect that application of $\tilde{\mathbf{C}}$ followed by application of $\tilde{\mathbf{C}}^{\mathbf{T}}$ should take us back to where we started, so the two matrices are the inverses of each other. Show this by calculating out $\tilde{\mathbf{C}}\tilde{\mathbf{C}}^{\mathbf{T}}$ and $\tilde{\mathbf{C}}^{\mathbf{T}}\tilde{\mathbf{C}}$ using matrix multiplication. Matrices whose transposes are their own inverses are called orthogonal matrices, and they are favored for ease of use.

Problem 10 Rotations About the $\hat{x}$ and $\hat{y}$ Axes

Go through the same projection process for computing the matrices for rotation about the other two axes. Assume rotation through a positive angle α about the $\hat{x}$ axis and derive the rotation matrix $\tilde{\mathbf{A}}$, then start over for a positive angle β about the $\hat{y}$ axis and derive the rotation matrix $\tilde{\mathbf{B}}$. You should find

$$\tilde{\mathbf{A}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix}, \tilde{\mathbf{B}} = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}. \tag{2.53}$$

We now have a method to arrive at a general answer to the question, "How do we perform the coordinate transformation?" The answer is that we repeat what we just did. The basis vectors are just vectors. Let's project them onto each other. There is no need to project the basis vectors within one set onto each other, for we know that the projections are either zero, if the two vectors are different, or one, if they are the same. This is because we have chosen orthonormal basis vectors within each system. The important projections are pairwise projections of a basis vector in one system onto a basis vector in the other. Projection simply involves taking a dot product.

If $\hat{x}, \hat{y}$, and $\hat{z}$ describe the system that we are transforming *from*, and $\hat{x}', \hat{y}'$, and $\hat{z}'$ describe the system that we are transforming *to*, then the transformation matrix is

$$\begin{pmatrix} \hat{x}'\cdot\hat{x} & \hat{x}'\cdot\hat{y} & \hat{x}'\cdot\hat{z} \\ \hat{y}'\cdot\hat{x} & \hat{y}'\cdot\hat{y} & \hat{y}'\cdot\hat{z} \\ \hat{z}'\cdot\hat{x} & \hat{z}'\cdot\hat{y} & \hat{z}'\cdot\hat{z} \end{pmatrix}. \tag{2.54}$$

2.5.3 More Challenging Transformations

There are two ways to perform more challenging transformations. One is stepwise successive rotations about two or three axes. The other is transformation in a single step using spherical trigonometry. We will use the single-step process first to make the promised transformation from Perifocal coordinates to G-E coordinates, then develop the stepwise rotations.

Single-Step Transformation from Perifocal to G-E Coordinates

The first part of this transformation is easy, for we simply apply the rule defined above. The basis vectors of Perifocal coordinates, the system from which we are transforming, are $\hat{P}, \hat{Q}$, and $\hat{W}$. The basis vectors of Geocentric-Equatorial coordinates, the system to which we are transforming, are $\hat{I}, \hat{J}$, and $\hat{K}$. The transformation matrix, called $\tilde{\mathbf{R}}$ in the text, is

$$\tilde{\mathbf{R}} = \begin{pmatrix} \hat{I}\cdot\hat{P} & \hat{I}\cdot\hat{Q} & \hat{I}\cdot\hat{W} \\ \hat{J}\cdot\hat{P} & \hat{J}\cdot\hat{Q} & \hat{J}\cdot\hat{W} \\ \hat{K}\cdot\hat{P} & \hat{K}\cdot\hat{Q} & \hat{K}\cdot\hat{W} \end{pmatrix}. \tag{2.55}$$

If only things could remain so easy.

Next comes the challenging part - to evaluate the dot products. Look at the challenge that we face. We must find the dot product of two vectors like $\hat{I}$ and $\hat{P}$, which means finding the cosine of angle between them. We don't know the angle, so we cannot immediately determine its cosine. We do know the angle from $\hat{I}$ to $\vec{n}$, the line of nodes, and we do know the angle from $\vec{n}$ to $\hat{P}$. It turns out that with this knowledge we can find the cosine of the angle between $\hat{I}$ and $\hat{P}$. To do this we need to do spherical trigonometry.

The material in this section is adapted from the long established and highly regarded

"Textbook on Spherical Astronomy" by W. M. Smart, sixth edition revised by R. M. Green, Cambridge University Press, Cambridge and New York, copyright 1977, reprinted 1986.

The first edition of this book was published in 1931, and a measure of its success is that it remains in print today. Following the convention in astrophysics, we will refer to this text as Smart (1977).

We are all familiar with plane trigonometry, which we have been using regularly in this class. This is the study of triangles in a plane. By extension, spherical trigonometry is the study of triangles on the surface of a sphere. Spherical trigonometry is essential to navigation, geodesy, and astronomy. Owing to these practical uses, it would have been studied by educated people in the time of Newton and Galileo, although it is not commonly studied today.

To get started we must have at our fingertips a result from plane trigonometry - the Law of Cosines. There is a version of this law in Euclid's Elements, Book II, Propositions 12 and 13, which is nearly 2500 years old. We will use a modern statement based on vectors that can be found on Wikipedia at

en.wikipedia.org/wiki/Law_of_cosines,

which we cite as

Wikipedia contributors, "Law of cosines," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Law_of_cosines&oldid=158080253 (accessed September 16, 2007).

We start with a triangle of sides $\vec{a}, \vec{b}$, and $\vec{c}$ such that

$$\vec{a} = \vec{b} - \vec{c}. \tag{2.56}$$

We square both sides using dot products to obtain

$$a^2 = b^2 + c^2 - 2\vec{b} \cdot \vec{c} = b^2 + c^2 - 2bc \cos\theta, \tag{2.57}$$

where a, b, and c are the magnitudes of the vectors and θ is the angle between $\vec{b}$ and $\vec{c}$. This result is so familiar that I am mildly surprised that it has such a grand name as Law of Cosines.

We will use this result to derive a related equation for spherical triangles. Our approach will be to construct plane triangles whose properties are easily related to those of spherical triangles. We will do this by intersecting a sphere with planes, similar to the way a plane intersects a cone to give conic sections. First we need a clear definition of a spherical triangle.

We choose to classify the intersection of a plane with a sphere according to whether the plane passes through the center of the sphere or not. If the plane does pass through the sphere's center then the resulting intersection with the sphere's surface is called a great circle. If it does not then the resulting intersection is called a small circle. In studying spherical triangles we limit ourselves to those triangles whose sides are arcs of great circles.

We choose a spherical triangle on the Celestial Sphere to analyze, and in our particular application the vertices of this triangle will be the points where unit vectors like $\hat{I}, \hat{J}, \hat{K}, \hat{P}, \hat{Q}$, and $\hat{W}$ intersect the Celestial Sphere. There is one other vector that we must include, and that is the node vector, $\vec{n}$. The line of nodes lies in both the fundamental plane and the orbital plane. The node vector is going to be the key to transforming from one to the other because it lies in both.

Following Smart (1977), and particularly Figure 3, label the vertices of this triangle A, B, and C, and label the arcs - segments of great circles - that define their opposite sides, BC, CA, and AB, as a, b, and c, respectively. For simplicity, call this triangle ABC. We will use three letters with no additional marks to indicate triangles, and three letters with a LaTeX math mode wide hat, $\widehat{ABC}$, to indicate the angle between segments AB and BC. The angle $\widehat{CAB}$ in a spherical triangle will also be called simply angle A.

Label the center of the sphere O. For convenience of thought we imagine the sphere to be in contact with a tangent plane, with vertex A being shared by the sphere and the plane at the tangent point. You may want to visualize a basketball resting on a gym floor, with A at the point of contact.

Obviously, the line from the center of the sphere, O, to point A intersects the plane at A. Construct the line from O to B, and extend it to its intersection with the plane at D. Similarly, construct the line from O to C and extend it to its intersection with the plane at E. These constructions allow us to define four plane triangles related to the spherical triangle. Triangle OAD includes the vertices A and B and the arc c. Triangle OAE includes the vertices A and C and the arc b. Triangle ODE includes the vertices B and C and the arc a. Triangle ADE is in the tangent plane, and its sides AD and AE are segments of lines that are tangent to the sphere and perpendicular to OA. This means that angles $\widehat{OAD}$ and $\widehat{OAE}$ are right angles, so the triangles OAD and OAE are right triangles. We also note that the angle A of the spherical triangle ABC is equal to angle $\widehat{DAE}$ of the plane triangle ADE. Now we are ready for the derivation, in which we will calculate the length of segment DE twice, once using the two triangles OAD and OAE, and once using the triangle DOE. We will equate the results to get our desired result.

Start with the plane triangle OAD. Its angle $\widehat{OAD}$ is a right angle. Angle $\widehat{AOD}$ in the plane triangle is equal to angle $\widehat{AOB}$ in the spherical triangle ABC. Both of these angles are equal to c. In the plane triangle OAD the side AD is opposite to angle c and side OA is adjacent, so $\frac{AD}{OA} = \tan c$, or

$$AD = OA \tan c, \tag{2.58}$$

similarly, OD is the hypotenuse, and $\frac{OA}{OD}$ is the cosine of c, so

$$OD = OA \sec c. \tag{2.59}$$

A similar analysis of plane triangle OAE and angle b gives

$$AE = OA \tan b, \tag{2.60}$$

and

$$OE = OA \sec b. \tag{2.61}$$

Triangle ADE is not necessarily a right triangle, so we analyze it using the Law of Cosines for a plane triangle derived previously to get

$$DE^2 = AD^2 + AE^2 - 2AD \cdot AE \cos \widehat{DAE}, \tag{2.62}$$

where DE, AD, and AE are the lengths of the respective sides. We have already established that $\widehat{DAE} = A$. Using this, and substituting for AD and AE in terms of OA and the tangents of b and c gives

$$DE^2 = OA^2\big(\tan^2 c + tan^2 b - 2\tan b \tan c \cos A\big). \tag{2.63}$$

The length DE is also a side of plane triangle DOE. Another application of the Law of Cosines in this plane triangle gives

$$DE^2 = OD^2 + OE^2 - 2OD \cdot OE \cos \widehat{DOE}. \tag{2.64}$$

Recognition that $\widehat{DOE} = \widehat{BOC} = a$ and substitution for OD and OE in terms of OA and secants of b and c gives

$$DE^2 = OA^2\big(\sec^2 c + \sec^2 b - 2\sec b \sec c \cos a\big). \tag{2.65}$$

Equating the two values of DE^2 gives

$$\sec^2 c + \sec^2 b - 2\sec b \sec c \cos a = \tan^2 c + \tan^2 b - 2\tan b \tan c \cos A. \tag{2.66}$$

We can remember, derive, or look up the trig identities

$$\sec^2 c = 1 + \tan^2 c, \sec^2 b = 1 + \tan^2 b, \tag{2.67}$$

and simplify after substitution to find the desired result,

$$\cos a = \cos b \cos c + \sin b \sin c \cos A. \tag{2.68}$$

This is the Law of Cosines for a spherical triangle, also called the Cosine Formula by Smart (1977). There are also two obviously related formulas

$$\cos b = \cos c \cos a + \sin c \sin a \cos B, \tag{2.69}$$

and

$$\cos c = \cos a \cos b + \sin a \sin b \cos C. \tag{2.70}$$

What have we accomplished? Let's make this concrete. Let the direction of OB be the direction of one known unit vector, such as $\hat{I}$, and the direction of OC be that of another known unit vector, such at $\hat{P}$. Let the node vector, $\vec{n}$, be in the direction from O to A. We will see that we know angle A of the spherical triangle, plus the angles b and c. Thus, we can determine the required dot product, which is $\cos a$. The other two formulas do the same thing for the angles b and c.

Let the directions be as described above: OB is in the direction of $\hat{I}$, OC is in the direction of $\hat{P}$, and OA is in the direction of $\vec{n}$. Then the angle from $\hat{I}$ to $\vec{n}$ in the fundamental plane is Ω, and the angle from $\vec{n}$ to $\hat{P}$ in the orbital plane is ω. The angle A is $180^o - i$ in degrees or $\pi - i$ in radians. We apply the Law of Cosines to the spherical triangle, where $\hat{I} \cdot \hat{P} = \cos a$ is desired, and where $b = \Omega, c = \omega$, and $A = \pi - i$. Then

$$\begin{aligned} \hat{I} \cdot \hat{P} &= \cos\Omega\cos\omega + \sin\Omega\sin\omega\cos(\pi - i) \\ &= \cos\Omega\cos\omega - \sin\Omega\sin\omega\cos i. \end{aligned} \tag{2.71}$$

We do the same thing for the other elements of the transformation matrix. Our text names the elements R_{ij}, where we just calculated $R_{11} = \hat{I} \cdot \hat{P}$. We will use both names for clarity. We will make repeated use of the trigonometric identities for the cosine and sine of the sum of angles,

$$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B, \tag{2.72}$$

and

$$\sin(A \pm B) = \sin A\cos B \pm \cos A\sin B. \tag{2.73}$$

Next, calculate $R_{12} = \hat{I} \cdot \hat{Q}$, still using $\vec{n}$ as the third vector. The angle between $\hat{I}$ and $\vec{n}$ in the fundamental plane remains Ω, but the angle between $\vec{n}$ and $\hat{Q}$ in the orbital plane is now $\omega + \frac{\pi}{2}$, and A remains $\pi - i$, so

$$\begin{aligned} R_{12} = \hat{I} \cdot \hat{Q} &= \cos\Omega\cos(\omega + \frac{\pi}{2}) + \sin\Omega\sin(\omega + \frac{\pi}{2})\cos(\pi - i) \\ &= \cos\Omega\big(\cos\omega\cos\frac{\pi}{2} - \sin\omega\sin\frac{\pi}{2}\big) - \sin\Omega\big(\sin\omega\cos\frac{\pi}{2} + \sin\frac{\pi}{2}\cos\omega\big)\cos i \\ &= -\cos\Omega\sin\omega - \sin\Omega\cos\omega\cos i. \end{aligned} \tag{2.74}$$

Next, calculate $R_{13} = \hat{I} \cdot \hat{W}$, continuing with $\vec{n}$ as the third vector. The angle between $\hat{I}$ and $\vec{n}$ in the fundamental plane is still Ω. The angle between $\vec{n}$ and $\hat{W}$ perpendicular to the orbital plane is $\frac{\pi}{2}$, and the spherical angle A is $\frac{\pi}{2} - i$, so

$$R_{13} = \hat{I} \cdot \hat{W} = \cos\Omega\cos\frac{\pi}{2} + \sin\Omega\sin\frac{\pi}{2}\cos(\frac{\pi}{2} - i) = \sin\Omega\sin i. \tag{2.75}$$

Next, calculate $R_{21} = \hat{J} \cdot \hat{P}$. The angle between $\hat{J}$ and $\vec{n}$ in the fundamental plane is $\frac{\pi}{2} - \Omega$. The angle between $\vec{n}$ and $\vec{P}$ in the orbital plane is ω. The spherical angle between the planes at $\vec{n}$ is i, so

$$\begin{aligned} R_{21} = \hat{J} \cdot \hat{P} = \cos(\frac{\pi}{2} - \Omega) \cos\omega + \sin(\frac{\pi}{2} - \Omega) \sin\omega \cos i \\ = \sin\Omega \cos\omega + \cos\Omega \sin\omega \cos i. \end{aligned} \tag{2.76}$$

Continue with $R_{22} = \hat{J} \cdot \hat{Q}$. The angle between $\hat{J}$ and n remains $\frac{\pi}{2} - \Omega$. The angle between $\vec{n}$ and $\hat{Q}$ in the orbital plane is $\omega + \frac{\pi}{2}$, and the spherical angle is i, so

$$\begin{aligned} R_{22} = \hat{J} \cdot \hat{Q} = \cos(\frac{\pi}{2} - \Omega) \cos(\omega + \frac{\pi}{2}) + \sin(\frac{\pi}{2} - \Omega) \sin(\omega + \frac{\pi}{2}) \cos i \\ = -\sin\Omega \sin\omega + \cos\Omega \cos\omega \cos i. \end{aligned} \tag{2.77}$$

Next, $R_{23} = \hat{J} \cdot \hat{W}$. The angle between $\vec{n}$ and $\hat{W}$ perpendicular to the orbital plane is $\frac{\pi}{2}$, and the spherical angle is $\frac{\pi}{2} + i$, so

$$\begin{aligned} R_{23} = \hat{J} \cdot \hat{W} = \cos(\frac{\pi}{2} - \Omega) \cos\frac{\pi}{2} + \sin(\frac{\pi}{2} - \Omega) \sin\frac{\pi}{2} \cos(\frac{\pi}{2} + i) \\ = -\cos\Omega \sin i. \end{aligned} \tag{2.78}$$

Next, $R_{31} = \hat{K} \cdot \hat{P}$. The angle between $\hat{K}$ and $\vec{n}$ perpendicular to the fundamental plane is $\frac{\pi}{2}$. The angle between $\vec{n}$ and $\hat{P}$ in the orbital plane is ω, and the spherical angle at $\vec{n}$ is $\frac{\pi}{2} - i$, so

$$R_{31} = \hat{K} \cdot \hat{P} = \cos\omega \cos\frac{\pi}{2} + \sin\omega \sin\frac{\pi}{2} \cos(\frac{\pi}{2} - i) = \sin\omega \sin i. \tag{2.79}$$

The last challenging calculation is $R_{32} = \hat{K} \cdot \hat{Q}$. The angle between $\vec{n}$ and $\hat{Q}$ in the orbital plane is $\omega + \frac{\pi}{2}$, and the spherical angle is $\frac{\pi}{2} - i$, so

$$\begin{aligned} R_{32} = \hat{K} \cdot \hat{Q} = \cos(\omega + \frac{\pi}{2}) \cos\frac{\pi}{2} + \sin(\omega + \frac{\pi}{2}) \sin\frac{\pi}{2} \cos(\frac{\pi}{2} - i) \\ = \cos\omega \sin i. \end{aligned} \tag{2.80}$$

The last calculation, $R_{33} = \hat{K} \cdot \hat{W}$, is easy because the angle between the unit vectors is i, and

$$R_{33} = \hat{K} \cdot \hat{W} = \cos i. \tag{2.81}$$

We now have the full transformation matrix.

Stepwise Successive Rotations

By being methodical it is always possible to align the axes of two coordinate sets by no ore than three individual rotations. A good example is transforming between the Topocentric-Horizon and Geocentric-Equatorial frames. The basis vectors of the Topocentric-Horizon (T-H) frame depend on the location of the topos, and point south, east, and upward toward the local zenith. They are called $\hat{S}, \hat{E}$, and $\hat{Z}$. (*Topos* is Greek for place.) The fundamental plane is that of the horizon. The basis vectors of the Geocentric-Equatorial frame, $\hat{I}, \hat{J}, \hat{K}$, are already familiar. Figure 2.8-4 can be used to visualize the relationship of the two frames. The transformation from G-E coordinates to T-H coordinates requires two types or operations: a shift of origin and rotations to align the coordinate axes. We will deal with the shift of origin in a later section of this chapter. We can align the axes by rotating the G-E basis vectors about $\hat{Z}$ through a positive angle θ until $\hat{I}$ and $\hat{J}$ are aligned with $\hat{S}$ and $\hat{E}$, respectively. Then we can rotate about $\hat{J}$ through a positive angle $\frac{\pi}{2} - L$ until $\hat{K}$ aligns with $\hat{Z}$. The rotation about $\hat{Z}$ is described by the matrix

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.82}$$

and the rotation about $\hat{J}$ is described by

$$\begin{pmatrix} \cos\left(\frac{\pi}{2} - L\right) & 0 & -\sin\left(\frac{\pi}{2} - L\right) \\ 0 & 1 & 0 \\ \sin\left(\frac{\pi}{2} - L\right) & 0 & \cos\left(\frac{\pi}{2} - L\right) \end{pmatrix} = \begin{pmatrix} \sin L & 0 & -\cos L \\ 0 & 1 & 0 \\ \cos L & 0 & \sin L \end{pmatrix}. \tag{2.83}$$

The order of the rotation matters: the $\hat{Z}$ rotation must be applied first. The result is

$$\begin{pmatrix} \sin L & 0 & -\cos L \\ 0 & 1 & 0 \\ \cos L & 0 & \sin L. \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sin L\cos\theta & \sin L\sin\theta & -\cos L \\ -\sin\theta & \cos\theta & 0 \\ \cos L\cos\theta & \cos L\sin\theta & \sin L \end{pmatrix}. \tag{2.84}$$

The text calls this matrix $\tilde{\mathbf{D}}$.

Problem 11 More on Orthogonal Matrices

We have seen that if a matrix is orthogonal then its transpose is equal to its inverse. Another way to define orthogonal matrices is that the elements of each row and column form a unit vector, and the vectors are orthonormal. Show that $\tilde{\mathbf{D}}$ has this property, and confirm that $\tilde{\mathbf{D}}^{\mathbf{T}} = \tilde{\mathbf{D}}^{\mathbf{-1}}$.

Owing to the rotation of the Earth the topos is also rotating, so θ is constantly increasing. We must account for this with precision appropriate to our goals.

2.6 Mechanics on the Rotating Earth

We live on the surface of the rotating Earth, and usually choose the reference frame of coordinates to be fixed to the Earth. This reference frame is accelerating, so it is not a true inertial frame. It may be close enough to inertial for some purposes, but not for orbital mechanics. We must learn how to keep time and to determine positions, velocities, and accelerations in this frame.

2.6.1 An Introduction to Time

The Earth's axial rotation and orbital motion, along with earthquakes and friction between the ocean waters and the ocean floor, all complicate the keeping of time.

Solar Time

Clocks were uncommon during the time of Galileo. He measured time in some of his experiments by using his pulse in place of a clock. In others he measured the collected volume of water from a flowing source as a measure of time. People's daily activities were, and largely continue to be, in accord with the apparent motion of the Sun. Apparent noon is marked by the presence of the Sun on the local meridian, and an apparent day defined as the time between successive apparent noons. Time kept in this way is called Apparent Solar Time or Local Apparent Time.

Ongoing, precise astronomical observations showed that days are not equal in length, that is, the time between successive meridian passages of the Sun changes slightly and smoothly throughout the year. This led to the definition of a fictitious Mean Sun that always produces days of exactly 24 hours or 86,400 seconds, so we could keep Mean Solar Time, or civil time, with a civil clock. This is the type of clock with which we are all familiar. The version of civil time kept by the *Bureau International des Poids et Mesures* and the Time Service Department of the U. S. Naval Observatory is called Universal Time, or UT. Please read about the BIPM at

www.bipm.org,

UT at

aa.usno.navy.mil/faq/docs/UT

and the Time Service Department at,

tycho.usno.navy.mil/time.html.

The difference between Mean Solar Time and Apparent Solar Time is called the Equation of Time, and a graphical plot of the Equation of Time is called an analemma. An analemma looks like a thin figure eight, and is often printed on globes. Please read about time at

www.mysundial.ca/tsp/time.html,

and view an amazing illustration of an analemma at

http://www.twanight.org/newTWAN/photos.asp?ID=3001422.

Sidereal Time

It also became apparent that the Sun moves relative to the stars, making a complete circuit in one year. The result is that we see different constellations during different times of the year, and since Newton's time we have taken the "fixed stars" to provide an inertial frame of reference. (In fact it is not inertial because of rotation of the Galaxy. None the less, the apparently fixed stars still provide a very useful reference frame called the Local Standard of Rest, or LSR, which could be used for navigation to the nearest stars.) Thus, it became possible to define the sidereal day as successive meridian passages of a star (from the Latin *sidus* or *sider*, meaning star), in addition to the definition of the solar day as successive meridian passages of the Sun. This led to the distinction between a solar day and a sidereal day, with the sidereal day being shorter by about four minutes, so that there are 366 sidereal days in a year. Marking the sidereal day required Sidereal Time and a sidereal clock running slightly faster than a civil clock.

Successful navigation required an agreed upon Prime Meridian for measuring angles and time - for a while each country had its own - and a similar reference point on the sky - the Vernal Equinox or First Point of Aries. Angles on the Celestial Sphere and on the Earth could be measured not only in degrees and radians, but also in time. Similarly, time could be measured as an angle. Thus, we use hours, minutes, and seconds, such that 360 degrees equal 24 hours, so 1 hour equals 15 degrees, 1 minute of time equals 15 minutes of arc, and 1 second of time equals 15 seconds of arc. Local sidereal time is zero hours, or $0^h, 00^m, 00^s$, when the First Point of Aries is on the local meridian, and the Local Sidereal Time (LST) advances as the amount of time since the First Point was on the meridian. Please read about Sidereal Time at

tycho.usno.navy.mil/sidereal.html.

Mathematically we have

$$\theta = \theta_g + \lambda_E. \tag{2.85}$$

Here the quantity θ is the angle from the stationary $\hat{I}$ axis to the meridian of the object being observed, and θ_g is the angle between $\hat{I}$ and Greenwich - the Greenwich sidereal time. The quantity λ_E is longitude measured eastward from Greenwich. The Greenwich sidereal time advances at the sidereal rate, which is 2π radians in 24 sidereal hours or $23^h 56^m 04.^s 09054 = 86164.09054$ seconds of solar time. This rate is
$\omega_E = 1.0027379093$ rev day^{-1} $= 7.292115856 \times 10^{-5}$ rad s^{-1},
and we can write

$$\theta = \theta_{g0} + \omega_E(t - t_0) + \lambda_E, \tag{2.86}$$

where ω_E is the angular speed of the Earth, θ_{go} is the sidereal time at Greenwich at some temporal reference, t_0, usually 0^h Universal Time on January 1 of the year in question, and t is a general time running at the sidereal rate. This provides the time dependence of θ that we need to convert between G-E and T-H coordinates.

Example 5 An Example of How to Calculate Local Sidereal Time

What was the LST at a station on the equator at 57.296 degrees west longitude at 06:00 Greenwich Mean Time (GMT) on 2 January, 1970? This example is adapted from one on page 106 of the text. The year was chosen by the authors because the book was written in about 1970.

Page 104 of the text contains data from the *American Ephemeris and Nautical Almanac* for 1970. It shows that at 00:00 hours on 1 January 1970 the Greenwich Sidereal Time (GST) was $6^h 40^m 55.^s 061$ or 1.74933340 radians. This is θ_{g0} for 1970. The top of page 104 contains data on how to calculate day numbers for any given year. 1 January of each year is counted as day 0 of that year, with the day changing from 0 to 1 at the completion of the day. Thus, 2 January is day 1. At 06:00 hours the fraction $6.00/24.00 = 0.25$ of a day has passed, so the time of observation at the station is 1.25 days. This provides us with the term $t - t_0$ in units of days. The station is at 57.296 west longitude, which is 1.000 radian west. Thus, $\lambda_E = -1.000$. The LST at the station is then

$$\theta = \theta_{g0} + \omega_E(t - t_0) + \lambda_E, \tag{2.87}$$

where we choose to measure all angles in radians. Thus

$$\theta = 1.74933340 + 1.0027379093 \times 2\pi \times 1.25 - 1.000 = 8.62481852. \tag{2.88}$$

In this case $t - t_0$ is greater than one day, so θ is greater than 2π radians. It is advantageous, and in many cases astronomically necessary, to subtract the whole number of days and specify the sidereal time as the remainder. This gives

$$\theta = 8.62481852 - 6.28318531 = 2.34163322. \tag{2.89}$$

This brings up the question of how many decimal places are necessary to keep in a calculation like this. There are $24 \times 60 \times 60 = 86400$ seconds in a day, so one second is 0.00001156 of a day. To keep accuracy to one second we will need at least 5 decimal places. To calculate the time, position, and rendezvous of satellites or missile interceptors it may be necessary to keep three or four more decimal places.

Dynamical Time

The study of dynamics requires Dynamical Time, which marches forward smoothly at one second per second and one day per day, with each second and each day of equal duration. This seems overly simple, but it is not. Tidal friction and earthquakes can cause the Earth's rotation to slow, requiring the occasional introduction of a leap second. Leap years have leap days. Calendars are reformed. These cannot cause the celestial bodies to leap forward or backward arbitrarily and instantaneously, hence the need for smoothly marching Dynamical Time.

A uniform march of days is tracked using Julian days and the Julian date. As I write on 2007 July 19 (the method preferred by the International Astronomical Union for writing calendar dates) it is Julian day 2454301, which specifies the day as a whole number, and Julian date 2454301.40444, which specifies the time of day as well. Please read about Julian dates at

aa.usno.navy.mil/data/docs/JulianDate.html,

and systems of time, including Dynamical Time, at

tycho.usno.navy.mil/systime.html.

2.6.2 Position, Velocity, and Acceleration

Let $\vec{\rho}$ be the position of a satellite or space object observed in the T-H coordinates of a radar site. Please keep in mind that these T-H coordinates are attached to the surface of the rotating Earth, so they are not inertial. We wish to transform this vector to G-E coordinates. For the moment, assume that G-E coordinates are inertial and that the Earth is spherical to keep things simple. Inertial implies that the center of the Earth is at rest or in uniform motion. We know this to be untrue, because the Earth is accelerating, but we make the assumption anyway because the acceleration is small. Let $\vec{R}$ be the position of the radar site relative to the origin of G-E coordinates. Please remember that G-E coordinates have their origin at the center of the Earth, but do not rotate with the Earth, so $\vec{R}$ is the vector from the center of the Earth to the rotating radar site. The position of the space object, $\vec{r}$, is

$$\vec{r} = \vec{R} + \vec{\rho}. \tag{2.90}$$

So far this is just a simple shift of origin that was mentioned, but not described, in the earlier section on stepwise rotations.

It is worth pointing out that $\vec{R}$ and $\vec{\rho}$ should be expressed in the same coordinates, either T-H or G-E in this case, before they are added. If the Earth is spherical then $\vec{R}$ is especially simple in T-H coordinates, for

$$\vec{R} = r_E \hat{Z}, \tag{2.91}$$

where r_E is the radius of the Earth.

To find the velocity it is tempting to simply take a time derivative and write

$$\dot{\vec{r}} = \dot{\vec{R}} + \dot{\vec{\rho}}, \tag{2.92}$$

so we need to know $\dot{\vec{R}}$ and $\dot{\vec{\rho}}$.

The Change in $\vec{R}$

The Earth's rotation means that $\vec{R}$ is time dependent, and we wish to know that dependence. We note that different points on the surface of the Earth are moving in different directions and at different speeds. On a spherical Earth each point traces out a daily circle, so the direction of motion of each point is constantly changing.

There is a simple way to describe the motion of each point on the ideal, spherical Earth with a single equation. We define the Earth's angular velocity vector, $\vec{\omega}_E$. Its magnitude is equal to ω_E, the sidereal rotation rate. A right-hand rule is used to determine its direction. When the fingers of the right hand curl in the direction of axial rotation the right thumb points in the direction of $\vec{\omega}_E$. The Earth rotates toward the east, so $\vec{\omega}_E$ is parallel to the Earth's rotation axis and in the direction from the South Pole to the North Pole. The velocity is then

$$\dot{\vec{R}} = \vec{\omega}_E \times \vec{R}, \tag{2.93}$$

and, identifying $\vec{v} = \dot{\vec{r}}$ as the velocity in the inertial G-E frame and $\vec{V} = \dot{\vec{R}}$ as the velocity of the non-inertial T-H frame at the radar site gives

$$\vec{v} = \dot{\vec{\rho}} + \vec{V} = \dot{\vec{\rho}} + \vec{\omega}_E \times \vec{R}. \tag{2.94}$$

The Change in $\vec{\rho}$

Figure 2.7-1 shows T-H coordinates and their basis vectors, $\hat{S}, \hat{E}$, and $\hat{Z}$ in the south, east, and zenith directions, respectively. The position vector of an object in these coordinates is $\vec{\rho}$. The direction of $\vec{\rho}$ is specified using elevation and azimuth, given the symbols $E\ell$ and Az. The elevation of an object is the angular distance from the fundamental or horizon plane to the object. The azimuth is the angle from north going eastward to the projection of $\vec{\rho}$ onto the horizontal plane. The magnitude of $\vec{\rho}$, ρ, can be provided in real time by the return signal received by the radar. The encoders in the mounts of radar sets and the Doppler shift of the returned signal can be set up to provide values of the six quantities $\rho, \dot{\rho}, E\ell, \dot{E\ell}, Az$, and $\dot{Az}$. These are used to express the position vector relative to the radar site as

$$\vec{\rho} = \rho_S \hat{S} + \rho_E \hat{E} + \rho_Z \hat{Z}, \tag{2.95}$$

where the figure can be used to derive the components of $\vec{\rho}$ as

$$\begin{aligned} \rho_S &= -\rho \cos E\ell \cos Az \\ \rho_E &= \rho \cos E\ell \sin Az \\ \rho_Z &= \rho \sin E\ell. \end{aligned} \tag{2.96}$$

The velocity vector relative to the radar site is not so simple as we might wish, because the basis vectors $\hat{S}, \hat{E}, \hat{Z}$ are fixed to the rotating Earth, so they are rotating as well. When we take the time derivative of $\vec{\rho}$ we get

$$\dot{\vec{\rho}} = \dot{\rho}_S \hat{S} + \dot{\rho}_E \hat{E} + \dot{\rho}_Z \hat{Z} + \rho_S \dot{\hat{S}} + \rho_E \dot{\hat{E}} + \rho_Z \dot{\hat{Z}}. \tag{2.97}$$

The first set of components of the velocity are

$$\begin{aligned} \dot{\rho}_S &= -\dot{\rho} \cos E\ell \cos Az + \rho \dot{E\ell} \sin E\ell \cos Az + \rho \dot{Az} \cos E\ell \sin Az \\ \dot{\rho}_E &= \dot{\rho} \cos E\ell \sin Az - \rho \dot{E\ell} \sin E\ell \sin Az + \rho \dot{Az} \cos E\ell \cos Az \\ \dot{\rho}_Z &= \dot{\rho} \sin E\ell + \rho \dot{E\ell} \cos E\ell. \end{aligned} \tag{2.98}$$

The second set can be found by treating the basis vectors as we would any other rotating vector, so

$$\dot{\hat{S}} = \vec{\omega}_E \times \hat{S}, \dot{\hat{E}} = \vec{\omega}_E \times \hat{E}, \dot{\hat{Z}} = \vec{\omega}_E \times \hat{Z}. \tag{2.99}$$

We then multiply each vector cross product by its multiplyer, and find that the sum simplifies to $\vec{\omega}_E \times \vec{\rho}$. The end result is

$$\dot{\vec{\rho}} = \dot{\rho}_S \hat{S} + \dot{\rho}_E \hat{E} + \dot{\rho}_Z \hat{Z} + \vec{\omega}_E \times \vec{\rho}. \tag{2.100}$$

The Result for Velocity

Adding the results for $\dot{\vec{R}}$ and $\dot{\vec{\rho}}$, we get

$$\begin{aligned} \dot{\vec{r}} &= \vec{\omega}_E \times \vec{R} + \dot{\rho}_S \hat{S} + \dot{\rho}_E \hat{E} + \dot{\rho}_Z \hat{Z} + \vec{\omega}_E \times \vec{\rho} \\ &= \vec{\omega}_E \times (\vec{R} + \vec{\rho}) + \dot{\rho}_S \hat{S} + \dot{\rho}_E \hat{E} + \dot{\rho}_Z \hat{Z} \\ &= \vec{\omega}_E \times \vec{r} + \dot{\rho}_S \hat{S} + \dot{\rho}_E \hat{E} + \dot{\rho}_Z \hat{Z}. \end{aligned} \tag{2.101}$$

Acceleration

To determine the acceleration we expect that we have to take another time derivative. Before we do this we have to be clear about the meaning of each term in the equations above, so we know just what derivatives to take. Workers at the radar site make the observation of $\vec{\rho}$ in their non-inertial T-H frame, and also determine $\dot{\rho}$ in that frame. Hypothetical workers in the inertial G-E frame determine $\vec{r}$ and $\vec{v}$. Both sets of workers agree on the values of $\vec{\omega}_E$ and $\vec{R}$. This is partly because they can do a simple experiment to determine who is in the inertial, or more nearly inertial frame, by reading a very accurate accelerometer that they each carry. The frame with the smallest acceleration is most nearly inertial.

When we take the next time derivative to relate the accelerations we must take all the derivatives in the same frame, and we know that we want to take them in the G-E frame, because that is most nearly inertial. Rather than do this by brute force we seek an easier way.

Consider a general vector $\vec{A}$ to be expressed in both G-E and T-H coordinates. Following earlier practice we write

$$\vec{A} = A_I\hat{I} + A_J\hat{J} + A_K\hat{K} = A_S\hat{S} + A_E\hat{E} + A_Z\hat{Z}, \tag{2.102}$$

where the basis vectors are already familiar. The time derivative of $\vec{A}$ is

$$\begin{aligned}\frac{d\vec{A}}{dt} &= \dot{A}_I\hat{I} + \dot{A}_J\hat{J} + \dot{A}_K\hat{K} + A_I\dot{\hat{I}} + A_J\dot{\hat{J}} + A_K\dot{\hat{K}} \\ &= \dot{A}_S\hat{S} + \dot{A}_E\hat{E} + \dot{A}_Z\hat{Z} + A_S\dot{\hat{S}} + A_E\dot{\hat{E}} + A_Z\dot{\hat{Z}}. \end{aligned}\tag{2.103}$$

We take the G-E system to be inertial, so its basis vectors are constant, but those of the T-H frame are not, so

$$\frac{d\vec{A}}{dt} = \dot{A}_I\hat{I}+\dot{A}_J\hat{J}+\dot{A}_K\hat{K} = \dot{A}_S\hat{S}+\dot{A}_E\hat{E}+\dot{A}_Z\hat{Z}+A_S\dot{\hat{S}}+A_E\dot{\hat{E}}+A_Z\dot{\hat{Z}}. \tag{2.104}$$

The time derivatives of the T-H basis vectors are given by the formula above, so

$$\dot{\hat{S}} = \vec{\omega}_E \times \hat{S}, \dot{\hat{E}} = \vec{\omega}_E \times \hat{E}, \dot{\hat{Z}} = \vec{\omega}_E \times \hat{Z}. \tag{2.105}$$

The time derivatives become

$$\begin{aligned}\frac{d\vec{A}}{dt} &= \dot{A}_I\hat{I} + \dot{A}_J\hat{J} + \dot{A}_K\hat{K} \\ &= \dot{A}_S\hat{S} + \dot{A}_E\hat{E} + \dot{A}_Z\hat{Z} + A_S\vec{\omega}_E \times \hat{S} + A_E\vec{\omega}_E \times \hat{E} + A_Z\vec{\omega}_E \times \hat{Z}. \end{aligned}\tag{2.106}$$

We identify

$$\frac{d\vec{A}}{dt}|_i = \dot{A}_I\hat{I} + \dot{A}_J\hat{J} + \dot{A}_K\hat{K}, \quad (2.107)$$

as the time derivative of $\vec{A}$ in the inertial G-E frame, and

$$\frac{d\vec{A}}{dt}|_n = \dot{A}_S\hat{S} + \dot{A}_E\hat{E} + \dot{A}_Z\hat{Z}, \quad (2.108)$$

as the time derivative of $\vec{A}$ in the non-inertial T-H frame, where the basis vectors $\hat{S}, \hat{E}, \hat{Z}$ appear fixed. Reorganizing the expressions for the rotating basis vectors gives

$$\begin{aligned} A_S\vec{\omega}_E \times \hat{S} + A_E\vec{\omega}_E \times \hat{E} + A_Z\vec{\omega}_E \times \hat{Z} &= \vec{\omega}_E \times A_S\hat{S} + \vec{\omega}_E \times A_E\hat{E} + \vec{\omega}_E \times A_Z\hat{Z} \\ &= \vec{\omega}_E \times \vec{A}. \quad (2.109) \end{aligned}$$

This allows us to write

$$\frac{d\vec{A}}{dt}|_i = \frac{d\vec{A}}{dt}|_n + \vec{\omega}_E \times \vec{A}. \quad (2.110)$$

We have assumed nothing about $\vec{A}$ except that it is a vector, to the result applies to all vectors. We apply it to the position vector, $\vec{r}$, written above as

$$\vec{r} = \vec{R} + \vec{\rho}, \quad (2.111)$$

to get

$$\frac{d\vec{r}}{dt}|_i = \frac{\vec{r}}{dt}|_n + \vec{\omega}_E \times \vec{r}, \quad (2.112)$$

or

$$\vec{v}_i = \vec{v}_n + \vec{\omega}_E \times \vec{r}, \quad (2.113)$$

where $\vec{v}_i$ is the velocity observed in the inertial R-E frame and $\vec{v}_n$ is the velocity observed in the non-inertial T-H frame. We can apply our rule again to $\vec{v}_i$ to obtain

$$\begin{aligned} \frac{d\vec{v}_i}{dt}|_i = \vec{a}|_i &= \frac{d\vec{v}_i}{dt}|_n + \vec{\omega}_E \times \vec{v}_i \\ &= \frac{d\vec{v}_n}{dt}|_n + \frac{d(\vec{\omega}_E \times \vec{r})}{dt}|_n + \vec{\omega}_E \times \vec{v}_n + \vec{\omega}_E \times (\vec{\omega}_E \times \vec{r}) \\ &= \vec{a}_n + \dot{\vec{\omega}}_E \times \vec{r} + \vec{\omega}_E \times \vec{v}_n + \vec{\omega}_e \times \vec{v}_n + \vec{\omega}_E \times (\vec{\omega}_E \times \vec{r}) \\ &= \vec{a}_n + \dot{\vec{\omega}}_E \times \vec{r} + 2\vec{\omega}_E \times \vec{v}_n + \vec{\omega}_E \times (\vec{\omega}_E \times \vec{r}). \quad (2.114) \end{aligned}$$

This is the correct relation among the accelerations in the inertial and non-inertial frames. If the mass of the moving object is constant then multiplying through by the mass gives the relation among the forces.

Note that because the rotating frame is not inertial it has three accelerations proportional to $\vec{\omega}_E$ that do not appear in the inertial frame. We note that the Earth's rotation rate is essentially constant, so $\dot{\vec{\omega}}_E$ may be taken to be zero. This is an approximation that ignores the slow change in ω_E due to friction between the ocean water and the ocean bottom, and ignores small but sudden changes in ω_E from earthquakes that cause changes in the Earth's moment of inertia. The other two terms cannot be neglected.

The term $\vec{\omega}_E \times (\vec{\omega}_E \times \vec{r})$ is the acceleration that keeps the rotating coordinate system moving about the center of the Earth, and is called the centrifugal acceleration. This is a general result that applies to all circular motion of constant speed, hence its association with a centrifuge. The term $2\vec{\omega}_E \times \vec{v}_n$ provides the new and interesting effects, and is called the Coriolis acceleration. It is present only when $\vec{v}_n$ is not equal to zero, or when there is motion of the space object as observed in the non-inertial T-H frame.

It is customary to solve for the acceleration in the non-inertial frame, $\vec{a}_n$, to get

$$\vec{a}_n = \vec{a}_i - 2\vec{\omega}_E \times \vec{v}_n - \vec{\omega}_E \times (\vec{\omega}_E \times \vec{r}). \tag{2.115}$$

In many cases the only force in the inertial frame is gravity, so we replace $\vec{a}_i$ with $\vec{a}_g$ to get

$$\vec{a}_n = \vec{a}_g - 2\vec{\omega}_E \times \vec{v}_n - \vec{\omega}_E \times (\vec{\omega}_E \times \vec{r}), \tag{2.116}$$

where $\vec{a}_g$ is not a constant, but changes with radius from the gravitating body as $\frac{1}{r^2}$.

2.7 The Ellipsoidal Earth

Up to this point we have assumed that the Earth is spherical. To avoid errors of several miles in locations on the surface of the Earth we must abandon the spherical model in favor of an ellipsoidal one. The location of a launch or tracking station is specified by station coordinates.

The Earth's surface is well modeled by an ellipsoid - the surface generated by rotating an ellipse. At 6378.145 km the Earth's equatorial axes are larger than its polar axis, at 6356.785 km. Such an ellipsoid is called an oblate spheroid, as opposed to a prolate spheroid, one whose equatorial axes are smaller than the polar axis. The ellipsoidal model of the Earth is a good approximation to mean sea level, and is a true equipotential surface of the Earth's gravitational field - a plumb bob would hang perpendicular to this surface at all locations if local mass concentrations are ignored. This surface is called the geoid.

2.7.1 The Measurement of Latitude

The oblate shape of the Earth does not complicate the determination of longitude, but it does latitude. We can define latitude as the angle between the equator and a line from the station through the center of the Earth or as the angle between the equator and a line through the station perpendicular to the geoid, as in the Figure 2.8-1. The first is called the geocentric latitude, L', and the second is called the geodetic latitude, L. They are distinct in the ellipsoidal model, but would be identical in a spherical one. The geodetic latitude is the basis for most charts and maps. If a latitude is given and its exact nature is unspecified it is likely to be the geodetic latitude. There is also an astronomical latitude, which is the angle between the equator and the local normal uncorrected for mass concentrations. There is very little difference between the astronomical latitude and the geodetic latitude.

2.7.2 Station Coordinates

We seek a method for specifying station coordinates using geodetic latitude. To do this we draw an elliptical cross section of the Earth through the station under consideration, and draw the circle that bounds that ellipse on an $x-z$ coordinate set. The ellipse has major and minor axes a_e and b_e, so it is written as

$$\frac{x_e^2}{a_e^2} + \frac{z_e^2}{b_e^2} = 1, \tag{2.117}$$

and the circle is written as equal axes taken to be the same as a_e, so

$$\frac{x_c^2}{a_e^2} + \frac{z_c^2}{a_e^2} = 1. \tag{2.118}$$

We solve both for z^2 to get

$$z_e^2 = b_e^2\big(1 - \frac{x_e^2}{a_e^2}\big), \tag{2.119}$$

and

$$z_c^2 = a_e^2\big(1 - \frac{x_c^2}{a_e^2}\big). \tag{2.120}$$

We recognize that along a vertical line $x_e = x_c$, so on that line the z coordinates of the ellipse and circle circle are related by

$$\frac{z_e}{z_c} = \pm\frac{b_e}{a_e}, \tag{2.121}$$

and we choose the plus sign for points above the x axis.

We introduce the reduced latitude, which is angle β in the figure. Then the x and z coordinates of a point on the ellipse are given by

$$x = a_e \cos\beta, z = \frac{b_e}{a_e} a_e \sin\beta = b_e \sin\beta. \tag{2.122}$$

For any ellipse

$$b_e = a_e\sqrt{1-e^2}, \tag{2.123}$$

so

$$z = a_e\sqrt{1-e^2}\sin\beta. \tag{2.124}$$

The goal now becomes to express β in terms of geodetic latitude, L. We do this using the line tangent to the ellipse at the location of the station. You may remember from earlier course work that the slope of this line is dz/dx and the slope of the line perpendicular to it is $-dx/dz$. The slope of the normal is just $\tan L$, so

$$\tan L = -\frac{dx}{dz}. \tag{2.125}$$

We can obtain the differentials from the coordinates themselves, so

$$dx = -a_e \sin\beta d\beta, dz = a_e\sqrt{1-e^2}\cos\beta d\beta, \tag{2.126}$$

and

$$\tan L = \frac{\tan\beta}{\sqrt{1-e^2}}, \tag{2.127}$$

or

$$\tan\beta = \sqrt{1-e^2}\tan L = \frac{\sqrt{1-e^2}\sin L}{\cos L}, \tag{2.128}$$

where the last step was made to facilitate the calculation of $\sin\beta$ and $\cos\beta$. To do this recognize that

$$\tan\beta = \frac{A}{B} \tag{2.129}$$

where

$$A = \sqrt{1 - e^2} \sin L, B = \cos L. \tag{2.130}$$

We reach our goal by finding

$$\sin \beta = \frac{A}{\sqrt{A^2 + B^2}} = \frac{\sqrt{1 - e^2} \sin L}{\sqrt{1 - e^2 \sin^2 L}}, \tag{2.131}$$

and

$$\cos \beta = \frac{B}{\sqrt{A^2 + B^2}} = \frac{\cos L}{\sqrt{1 - e^2 \sin^2 L}}. \tag{2.132}$$

We can now write the station coordinates in terms of the geodetic latitude, L, as

$$x = \frac{a_e \cos L}{\sqrt{1 - e^2 \sin^2 L}}, \tag{2.133}$$

and

$$z = \frac{a_e(1 - e^2) \sin L}{\sqrt{1 - e^2 \sin^2 L}}. \tag{2.134}$$

Now that we have accomplished this we must realize that we need to be able to express the coordinates of objects above the Earth as well. An object with elevation or height H above the Earth on the line normal to the ellipsoid has additional displacements

$$\Delta x = H \cos L, \Delta y = H \sin L. \tag{2.135}$$

We add these to the previous results to get

$$x = \left[\frac{a_e}{\sqrt{1 - e^2 \sin^2 L}} + H \right] \cos L, \tag{2.136}$$

and

$$z = \left[\frac{a_e(1 - e^2)}{\sqrt{1 - e^2 \sin^2 L}} + H \right] \sin L. \tag{2.137}$$

This expresses two of the three station coordinates in terms of the geodetic latitude, L, the equatorial radius of the Earth, a_e, and the height above the geoid, H. The third station coordinate is the local sidereal time, which is longitude of the station east of Greenwich plus the Greenwich sidereal time. We know this to be

$$\theta = \theta_g + \lambda_E. \tag{2.138}$$

The vector from the center of the Earth to the station in $\hat{I}, \hat{J}, \hat{K}$ coordinates, with allowance for the ellipsoidal Earth, is

$$\vec{R} = x\cos\theta\hat{I} + x\sin\theta\hat{J} + z\hat{K}. \tag{2.139}$$

This equation is used when the latitude is far to the north or south, or when high precision positions are need, or both. An example would be when launching a missile interceptor from Alaska.

Let's now use this in an example.

Example 6 Calculating $\vec{R}$ for a Radar Station

What is the position in G-E coordinates of a radar site on the equator at 57.296 degrees west longitude at 06:00 GMT on 2 January, 1970? What is the position of a point 6.378 km above the radar site? This is an adaptation of the example on page 106 of our text.

This should seem familiar, because we have already dealt with the calculation of local sidereal time at this radar site for the same data. We found that the local sidereal time since the beginning of 1970 was 8.62481852 radians.

We know that the radar station is on the surface of the Earth at the equator, so $x = 1.00$ and $z = 0$. We can also show this from the formulas for x and z, for if $L = 0$ then $\sin L = 0$, $\cos L = 1$, $x = a_e = 1$ DU, and $z = 0$. Then

$$\begin{aligned}\vec{R} &= x\cos\theta\hat{I} + x\sin\theta\hat{J} + z\hat{K} \\ &= 1DU\cos 8.6248\hat{I} + 1DU\sin 8.6248\hat{J} + 0\hat{K} \\ &= -0.697DU\hat{I} + 0.717DU\hat{J}.\end{aligned} \tag{2.140}$$

Note that 6.378 km is 0.001 of the equatorial radius of the Earth, so the point above the radar site is at $x = 1.001$ DU, $z = 0$. Keeping three significant figures, this causes a very small change in $\vec{R}$, to

$$\vec{R} = -0.697DU\hat{I} + 0.718DU\hat{J}. \tag{2.141}$$

These positions are adequate for most mission planning, but would not be adequate for a missile interception, where it is desirable to know positions to sub-meter accuracy.

Example 7 Putting It All Together - The Value of a Single Radar Observation

This example is the culmination of all the work that we have done in the first half of the semester. It is an adaptation and expansion of the example that starts on page 107 of our text.

The Big Picture - If we want to know what a space object is doing right now then we need to specify six quantities: the three components of $\vec{r}_{IJK}$ and the three quantities of $\vec{v}_{IJK}$. We need this information in an inertial frame, and the $\hat{I}, \hat{J}, \hat{K}$ frame is the most familiar and convenient. If we want to know what that space object will do in the foreseeable future then we need to know its orbital elements, and we know how to calculate them from $\vec{r}$ and $\vec{v}$.

The Details - At 06:00 Greenwich Sidereal Time a tracking station at latitude 60^o north and longitude 150^o west detects a space object and provides the following data:

slant range $= \rho = 0.4DU$
azimuth $= Az = 90^o$
elevation $= E\ell = 30^o$
range rate $= \dot{\rho} = 0$
azimuth rate $= \dot{A}z = 10$ rad TU^{-1}
elevation rate $= \dot{E}\ell = 5$ rad TU^{-1}.

The Problem - Find $\vec{\rho}, \dot{\vec{\rho}}, \vec{r}$, and $\vec{v}$, then use $\vec{r}$ and $\vec{v}$ to find the orbital elements. Is this space object a threat?

Finding $\vec{r}$ and $\vec{v}$ - The data are given in degrees, so it is going to be convenient to work this example in degrees. We can easily find the components of $\vec{\rho}$ and $\dot{\vec{\rho}}$,

$$\begin{aligned}
\rho_S &= -0.4DU \cos 30^o \cos 90^o = 0 \\
\rho_E &= 0.4DU \cos 30^o \sin 90^o = 0.2\sqrt{3} = 0.346DU \\
\rho_Z &= 0.4DU \sin 30^o = 0.2DU \\
\dot{\rho}_S &= (0.4DU)(10rad/TU) \cos 30^o \sin 90^o = 2\sqrt{3} = 3.46DUTU^{-1} \\
\dot{\rho}_E &= -(0.4DU)(5rad/TU) \sin 30^o \sin 90^o = -1.0DUTU^{-1} \\
\dot{\rho}_Z &= (0.4DU)(5rad/TU) \cos 30^o = \sqrt{3} = 1.73DUTU^{-1}. \quad (2.142)
\end{aligned}$$

These are easily used to give

$$\vec{\rho} = (0.2\sqrt{3}\hat{E} + 0.2\hat{Z})DU = (0.346\hat{E} + 0.2\hat{Z})DU, \quad (2.143)$$

and

$$\dot{\vec{\rho}} = (2\sqrt{3}\hat{S} + -1.0\hat{E} + \sqrt{3}\hat{Z}) = (3.46\hat{S} - 1.0\hat{E} + 1.73\hat{Z})DUTU^{-1}. \quad (2.144)$$

We now have a choice. If we need high accuracy we must rotate ρ to the $\hat{I}, \hat{J}, \hat{K}$ system and then add $\vec{R}$ calculated in the $\hat{I}, \hat{J}, \hat{K}$ system with the formulas for the ellipsoidal Earth. If we do not need high accuracy we can assume

$$\vec{R} = 1\hat{Z}DU \quad (2.145)$$

in the $\hat{S}, \hat{E}, \hat{Z}$ system, add that value to $\vec{\rho}$ in the $\hat{S}, \hat{E}, \hat{Z}$ system, and then rotate the sum to the $\hat{I}, \hat{J}, \hat{K}$ system. That is, using the rotation matrix $\tilde{\mathbf{D}}^{\mathbf{-1}}$ from earlier in this chapter, we can either calculate

$$\vec{r}_{IJK} = \tilde{\mathbf{D}}^{\mathbf{-1}}\vec{\rho}_{SEZ} + \vec{R}_{IJK} \quad (2.146)$$

when high accuracy is needed, or

$$\vec{r}_{IJK} = \tilde{\mathbf{D}}^{\mathbf{-1}}(\vec{\rho}_{SEZ} + 1\hat{Z}) = \tilde{\mathbf{D}}^{\mathbf{-1}}\vec{\rho}_{SEZ} + \tilde{\mathbf{D}}^{\mathbf{-1}}\hat{Z} \quad (2.147)$$

when high accuracy is not needed. The difference is between the two approaches occurs because

$$\tilde{\mathbf{D}}^{\mathbf{-1}}\hat{Z} \neq \vec{R}_{IJK}, \quad (2.148)$$

although the difference is usually very small. We note in passing that taking $\vec{R} = 1\hat{Z}$ is the same as taking $e = 0$ in the equations for x and z for the ellipsoidal Earth.

Let's first assume that we do not need high accuracy. Then

$$\vec{r}_{SEZ} = \vec{\rho} + \hat{Z} = (0.2\sqrt{3}\hat{E} + 1.2\hat{Z})DU = 0.346\hat{E}DU + 1.2\hat{Z}DU. \quad (2.149)$$

To calculate the rotation matrix we need the local sidereal time. The tracking computer has conveniently provided us with the Greenwich sidereal time, and we know

$$\theta = \theta_g + \lambda_E. \quad (2.150)$$

The station is at 150^o west longitude, so, still working in degrees, $\lambda_E = -150^o$, and $\theta_g = 06:00/24:00$ revolution, or 90^o, and

$$\theta = 90^o - 150^o = -60^o. \quad (2.151)$$

The rotation matrix that we need, to transform from $\hat{S}, \hat{E}, \hat{Z}$ coordinates to $\hat{I}, \hat{J}, \hat{K}$ coordinates, is called $\tilde{\mathbf{D}}^{-1}$ in the text, and is given by

$$\tilde{\mathbf{D}}^{-1} = \begin{pmatrix} \sin L \cos\theta & -\sin\theta & \cos L \cos\theta \\ \sin L \sin\theta & \cos\theta & \cos L \sin\theta \\ -\cos L & 0 & \sin L \end{pmatrix}. \tag{2.152}$$

Plugging in the angles gives

$$\tilde{\mathbf{D}}^{-1} = \begin{pmatrix} \frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{2} & \frac{1}{4} \\ -\frac{3}{4} & \frac{1}{2} & -\frac{\sqrt{3}}{4} \\ -\frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \tag{2.153}$$

Our desired result is

$$\begin{aligned} \vec{r}_{IJK} = \tilde{\mathbf{D}}^{-1}\vec{r}_{SEZ} &= \begin{pmatrix} \frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{2} & \frac{1}{4} \\ -\frac{3}{4} & \frac{1}{2} & -\frac{\sqrt{3}}{4} \\ -\frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 0 \\ 0.2\sqrt{3} \\ 1.2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{3}{10} + \frac{3}{10} \\ \frac{\sqrt{3}}{10} - 0.3\sqrt{3} \\ 0.6\sqrt{3} \end{pmatrix} = \begin{pmatrix} 0.6 \\ -0.2\sqrt{3} \\ 0.6\sqrt{3} \end{pmatrix} \end{aligned} \tag{2.154}$$

Keeping five significant figures gives

$$\vec{r}_{IJK} = (0.60000\hat{I} - 0.34641\hat{J} + 1.0392\hat{K})DU. \tag{2.155}$$

If we do need high accuracy then we first rotate $\vec{\rho}$ to the $\hat{I}, \hat{J}, \hat{K}$ system to get

$$\vec{\rho}_{IJK} = \tilde{\mathbf{D}}^{-1} \begin{pmatrix} 0 \\ 0.2\sqrt{3} \\ 0.2 \end{pmatrix} = \begin{pmatrix} \frac{3}{10} + \frac{1}{20} \\ \frac{\sqrt{3}}{10} - \frac{\sqrt{3}}{20} \\ \frac{\sqrt{3}}{10} \end{pmatrix} = \begin{pmatrix} 0.35 \\ \frac{\sqrt{3}}{20} \\ \frac{\sqrt{3}}{10} \end{pmatrix} \tag{2.156}$$

Then we calculate $\vec{R}_{IJK}$ from

$$x = \frac{a_E}{\sqrt{1 - e^2 \sin^2 L}} \cos L \tag{2.157}$$

and

$$z = \frac{a_E(1 - e^2)}{\sqrt{1 - e^2 \sin^2 L}} \sin L \tag{2.158}$$

Owing to the terms in e^2 these cannot easily be kept in closed form. Working in canonical units, so $a_E = 1DU$, they give

$$x = 0.5012599DU \tag{2.159}$$

and

$$z = 0.8623955DU. \tag{2.160}$$

Then $\vec{R}_{IJK}$ is calculated from

$$\vec{R} = x\cos\theta\hat{I} + x\sin\theta\hat{J} + z\hat{K} \tag{2.161}$$

to give

$$\vec{R} = (0.2506299\hat{I} - 0.4341038\hat{J} + 0.8623995\hat{K})DU. \tag{2.162}$$

Adding as required, and keeping five significant figures gives

$$\vec{r}_{IJK} = \vec{\rho}_{IJK} + \vec{R}_{IJK} = (0.60069\hat{I} - 0.34750\hat{J} + 1.0356\hat{K})DU, \tag{2.163}$$

so the two approaches agree to about three significant figures. Remember that we have worked in canonical units, so to convert to kilometers each value must be multiplied by 6378.145. Then a difference of one unit in the fourth decimal place makes a difference of 0.6378 kilometers. That error is negligible for initial mission planning, but not for landing, interception, or rendezvous.

We still have to make the conversion from $\dot{\vec{\rho}}_{SEZ}$ to $\vec{v}_{IJK}$, using

$$\vec{v}_{IJK} = \tilde{\mathbf{D}}^{-1}\dot{\vec{\rho}}_{SEZ} + \vec{\omega}_E \times \vec{r}, \tag{2.164}$$

and, again, we have a choice. If we don't need high accuracy we can assume that the Earth is spherical so $\vec{R} = 1\hat{Z}$, or we can demand high accuracy and calculate $\vec{R}$ using the formulas for the ellipsoidal Earth. It is worth pointing out that we are dealing with velocity vectors, which are calculated from the difference of two position vectors divided by a time interval. Subtracting the positions takes the choice of origin out of the problem, so no origin shift is required.

First, the easy part.

$$\dot{\vec{\rho}}_{IJK} = \tilde{\mathbf{D}}^{-1}\dot{\vec{\rho}}_{SEZ} = \begin{pmatrix} \frac{3}{2} - \frac{\sqrt{3}}{4} \\ \frac{3\sqrt{3}}{2} - \frac{5}{4} \\ -\sqrt{3} + \frac{3}{2} \end{pmatrix}. \tag{2.165}$$

Keeping five significant figures gives

$$\dot{\vec{\rho}}_{IJK} = (1.0670\hat{I} - 3.8481\hat{J} - 0.23205\hat{K})DU. \tag{2.166}$$

The rotation vector, $\vec{\omega}_E$ has direction $\hat{K}$ and magnitude 7.29211×10^{-5} when expressed in radians per second. We are working in canonical units, so we need units of radians per TU. We multiply by 806.811 seconds per TU to get

$$\omega_E = 5.8833 \times 10^{-2} \tag{2.167}$$

in units of radians per TU. If we don't need high accuracy then

$$\vec{r}_{IJK} = (0.6\hat{I} - 0.2\sqrt{3}\hat{J} + 0.6\sqrt{3}\hat{K})DU, \tag{2.168}$$

and

$$\vec{\omega}_E \times \vec{r}_{IJK} = (0.2\sqrt{3}\omega_E\hat{I} + 0.6\omega_E\hat{J})DUTU^{-1} = (0.02038\hat{I} + 0.03530\hat{J})DUTU^{-1}. \tag{2.169}$$

Keeping four decimal places gives

$$\vec{v}_{IJK} = (1.0874\hat{I} - 3.8128\hat{J} - 0.2321\hat{K})DUTU^{-1}. \tag{2.170}$$

If we do want high accuracy then

$$\vec{r}_{IJK} = (0.60069\hat{I} - 0.34750\hat{J} + 1.0356\hat{K})DU, \tag{2.171}$$

and

$$\vec{\omega}_E \times \vec{r}_{IJK} = (2.0444 \times 10^{-2}\hat{I} + 3.5340 \times 10^{-2}\hat{J})DUTU^{-1}. \tag{2.172}$$

Adding and keeping four decimal places gives

$$\vec{v}_{IJK} = (1.0874\hat{I} - 3.8128\hat{J} - 0.23205\hat{K})DUTU^{-1}, \tag{2.173}$$

so the difference between the two approaches is insignificant at this level of accuracy. This is because the two values of $\vec{r}_{IJK}$ do not differ until the fourth decimal place and the magnitude of ω_E is significantly smaller than one.

The Orbital Elements - We need to calculate the three vectors $\vec{h}, \vec{n}$, and $\vec{e}$. Let's keep three significant figures in each vector component, so

$$\vec{r}_{IJK} = (0.600\hat{I} - 0.346\hat{J} + 1.04\hat{K})DU, \tag{2.174}$$

and

$$\vec{v}_{IJK} = (1.09\hat{I} - 3.38\hat{J} - 0.232\hat{K})DUTU^{-1}. \tag{2.175}$$

First, find the specific angular momentum,

$$\vec{h} = \vec{r} \times \vec{v} = \begin{vmatrix} \hat{I} & \hat{J} & \hat{K} \\ 0.600 & -0.346 & 1.04 \\ 1.09 & -3.83 & -0.232 \end{vmatrix}. \tag{2.176}$$

This gives

$$\vec{h} = (4.06\hat{I} + 1.27\hat{J} - 1.92\hat{K})DU^2TU^{-1}. \tag{2.177}$$

Before using this value of $\vec{h}$ to calculate the orbital elements, let's test to see if we have made any mistakes. Calculate

$$\vec{h} \cdot \vec{r} = -2.2 \times 10^{-4} \simeq 0, \tag{2.178}$$

and

$$\vec{h} \cdot \vec{v} = 6.74 \times 10^{-3} \simeq 0, \tag{2.179}$$

as expected, because $\vec{h}$ is perpendicular to both $\vec{r}$ and $\vec{v}$ by definition of the cross product.

Squaring the components gives $h^2 = 21.8$, or $h = 4.67$, and

$$p = \frac{h^2}{\mu} = 21.8DU. \tag{2.180}$$

This also allows us to calculate

$$\cos i = \frac{h_K}{h} = \frac{-1.92}{4.67}. \tag{2.181}$$

We know that i is always less than 180^o, so

$$i = 114.^o3. \tag{2.182}$$

Next let's find the node vector, $\vec{n}$,

$$\vec{n} = \hat{K} \times \vec{h} = \begin{vmatrix} \hat{I} & \hat{J} & \hat{K} \\ 0 & 0 & 1 \\ 4.06 & 1.27 & -1.92 \end{vmatrix}. \tag{2.183}$$

This gives

$$\vec{n} = -1.27\hat{I} + 4.06\hat{J}. \tag{2.184}$$

We can check this result by confirming that $\vec{n} \cdot \hat{K} = 0$ and $\vec{n} \cdot \vec{h} = 0$, which they do.

It is worth calculating $n^2 = 18.1$, and $n = 4.25$. We can now use $\vec{n}$ to find Ω,

$$\cos \Omega = \frac{n_I}{n} = \frac{-1.27}{4.25}. \tag{2.185}$$

We know that Ω is always less than 180^o, so

$$\Omega = 107^o.4 \tag{2.186}$$

Next we need the eccentricity vector, $\vec{e}$, given by

$$\vec{e} = \frac{1}{\mu}\left[\left(v^2 - \frac{\mu}{r}\right)\vec{r} - (\vec{r} \cdot \vec{v})\vec{v}\right]. \tag{2.187}$$

For this we need $v^2 = 15.9$, $v = 3.99$, $r^2 = 1.56$, $r = 1.25$, and $\vec{r} \cdot \vec{v} = 1.74$. This already provides a good bit of information about the orbit, for

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} \tag{2.188}$$

is now easy to apply. Remembering that in canonical units $\mu = 1$, it is easy to see that $\mathcal{E}$ is positive, so the orbit must be hyperbolic. The eccentricity vector becomes

$$\vec{e} = 15.1\vec{r} - 1.74\vec{v}. \tag{2.189}$$

This gives

$$\vec{e} = 7.16\hat{I} + 1.44\hat{J} + 16.1\hat{K}, \tag{2.190}$$

from which we get $e^2 = 312$ and $e = 17.7$. This confirms that the orbit is hyperbolic.

Knowing $\vec{e}$ we can now calculate ω from

$$\cos \omega = \frac{\vec{n} \cdot \vec{e}}{ne} = \frac{-1.27 \cdot 7.16 + 4.06 \cdot 1.44}{4.25 \cdot 17.7}. \tag{2.191}$$

We know that $e_k > 0$, so $\omega < 180^o$, and

$$\omega = 92^o.5. \tag{2.192}$$

Next we calculate

$$\cos \nu_0 = \frac{\vec{e} \cdot \vec{r}}{er} = \frac{0.6 \cdot 7.16 - 0.356 \cdot 1.44 + 1.04 \cdot 16.1}{17.7 \cdot 1.25}. \tag{2.193}$$

We know that $\vec{r} \cdot \vec{v} = 1.74 > 0$, so $\nu_0 < 180^o$, and

$$\nu_0 = 21^o.8. \tag{2.194}$$

We have calculated six orbital elements, so our work is complete, but we can calculate u_0 as a check of our work, for if we have done it correctly then

$$\omega + \nu_0 = u_0. \tag{2.195}$$

We know

$$\cos u_0 = \frac{\vec{n} \cdot \vec{r}}{nr} = \frac{-1.27 \cdot 0.6 + 4.06 \cdot -0.346}{4.25 \cdot 1.25}. \tag{2.196}$$

We see that $r_K > 0$, so $u_0 < 180^o$, and

$$u_0 = 114^o.1. \tag{2.197}$$

We apply the check,

$$\omega + \nu_0 = 92^o.5 + 21^o.8 = 114^o.3 \simeq 114^o.1, \tag{2.198}$$

so we have the orbital elements of the hyperbolic orbit.

Is the object a threat? We know

$$r = \frac{p}{1 + e\cos\nu}. \tag{2.199}$$

Is there a value of ν that allows $r = 1$? If so, then the object can hit the Earth. Rearranging gives

$$\cos\nu = \frac{1}{e}\left(\frac{p}{r} - 1\right). \tag{2.200}$$

Plugging in, using the condition $r = 1$, gives

$$\frac{1}{17.7}\left(\frac{21.8}{1} - 1\right) = 1.18 > 1. \tag{2.201}$$

There is no value of ν that gives $\cos\nu = 1.18$, so the object cannot hit the surface of the Earth, and it is not a threat.

The example definitely puts to use almost everything that we have done so far. It also shows that the most useful implementation of this work will be on a computer, so that the position, velocity, orbital elements, and threat status can be known immediately.

2.8 The Ground Track of a Satellite

We now have the ability to describe a satellite in several coordinate systems and to transform the description from one system to another. The next step is to describe the ground track of a satellite. First, let's define ground track. In the two-body limit the orbit of an Earth satellite is confined to a plane and the center of the Earth is located at one focus of the orbit. We consider the line joining the satellite to the center of the Earth, and in particular consider the point on this line that intersects the surface of the Earth. The motion of this point traces the ground track of the satellite. In the ideal case of a spherical Earth the ground track would be a great circle on the Earth's surface. If the ideal, spherical Earth did not rotate on its axis then the orbit would be closed and would repeat itself over and over, so the ground track would always be the same great circle. The Earth's rotation and departures from the ideal Earth and the ideal two-body problem cause the ground track to evolve with time.

The purpose or mission of the satellite determines the nature and importance of its ground track. A satellite surveying the Earth's surface must be able to point its cameras at the desired targets. A communications satellite must have line of sight to the stations that need to communicate. Satellites that need control commands in real time or that report data back to stations on the Earth must pass over transmitting and receiving stations. Manned satellites must pass over safe landing areas with reasonable frequency.

The eastward rotation of the Earth causes the ground track of low satellites to move westward. You are probably already familiar with this from watching news reporting of space missions.

2.8.1 Launch Site, Launch Azimuth, and Orbital Inclination

Figure 2.15-2 shows a satellite launch site at latitude L_o, illustrates the orbital inclination, i, and defines the launch azimuth, labeled β_o. The angles can be used to draw a spherical triangle on the surface of the Earth using the equator and the meridian of the launch site. We can use the Cosine Formula for a spherical triangle derived earlier to write

$$\cos i = -\cos B \cos \beta_o + \sin B \sin \beta_o \cos L_o, \tag{2.202}$$

where B is the angle at vertex B, which is the angle between the equator and the meridian. This angle is clearly 90 degrees or $\frac{\pi}{2}$ radians, so the equation simplifies to

$$\cos i = \sin \beta_o \cos L_o. \tag{2.203}$$

The simplicity of this equation disguises its importance. Suppose that we wish to minimize i - to get it equal to zero. The purpose of this would be to launch a satellite directly into equatorial orbit. Minimizing i maximizes $\cos i$ owing to the nature of the cosine function. Maximizing $\cos i$ means that $\sin \beta_o$ must be equal to one, so β_o must be 90 degrees or $\frac{\pi}{2}$ radians. Thus,

$$\cos i_{min} = \cos L_o, \tag{2.204}$$

and $i_{min} = L_o$. The latitude of the launch site sets the minimum of the orbital inclination. Think about why our most famous launch facility is in Florida, at a latitude of about 28.5 degrees. This allows launching directly into orbits with inclinations in the range $28.5 \leq i \leq 90$. Orbital missions requiring lower inclinations will start at an initial orbit at an inclination of about 28.5 degrees, followed by a subsequent change-of-plane orbital maneuver, described in the next chapter, to reach the desired inclination. This puts far northern countries at a disadvantage, for change-of-plane maneuvers are costly of fuel. This is one of the reasons why the Russian space program is notable for its large and powerful rockets.

While we are on the topic, there is another reason why Florida is a desirable launch site and why many space missions favor direct, or eastward, orbits. The Earth rotates to the east, so the launch site has a velocity in the direction of the eastward orbit. The magnitude of this velocity is $\omega_E r_E \cos L_o$, where ω_E is the angular velocity of the Earth, r_E is the radius of the Earth, and $\cos L_o$ is the latitude of the launch site. This velocity is largest at the equator and zero at the poles. The equatorial velocity is thus

$$(7.29 \times 10^5 rads^{-1})(6.378 \times 10^6 m) = 4.650 \times 10^2 ms^{-1} = 5.882 \times 10^{-2} DUTU^{-1}, \tag{2.205}$$

a small, but helpful advantage. Remember that the speed required for the Earth reference orbit is $1 DUTU^{-1}$, so at the equator the Earth's rotation provides 5.88 percent of the needed speed for the reference orbit. A circular orbit at an altitude of $100 nmi$ requires a speed of

$$v_{cs} = \sqrt{\frac{\mu}{r_{cs}}} = 0.9858 DUTU^{-1}, \tag{2.206}$$

just a bit slower than the reference orbit, but this makes the eastward motion of the launch site a little more helpful.

Note that if a westward or retrograde orbit is required then fuel must be burned to overcome the eastward motion, then a similar amount of fuel must be burned to achieve the same speed of westward motion. Thus, the advantage of the eastward orbit is doubled.

Problem 12 Locations of the World's Spaceports

Compare the eastward motions and relative advantages of the European Space Agency's spaceport at Kourou, French Guiana, at 5 degrees north latitude, Cape Canaveral at 28.5 degrees north latitude, White Sands Missile Range at 32.3 degrees north latitude, Vandenberg Air Force Base at 34.4 degrees north latitude, Baikonur Cosmodrome at 45.9 degrees north latitude, and Plesetsk Cosmodrome at 62.8 degrees north latitude.

The web site www.spacetoday.org/Rockets/Spaceports/LaunchSites.html contains a list of worldwide launch sites and their locations.

Chapter 3

Real Orbits and Orbital Maneuvers

3.1 Some Types of Orbits

We already know how to classify conic-section orbits by their shape, eccentricity, specific angular momentum, and specific mechanical energy. This is just the start. We now wish to classify orbits by their purpose and use.

3.1.1 Classification of Earth Orbit by Altitude

Low Earth Orbit or LEO

Figure 3.1-1 in the text makes it clear that an Earth orbit that is too low will decay because of the drag caused by the tenuous residual atmosphere. A circular orbit with an altitude of 50 miles decays in about an hour. Obviously, higher orbits take longer to decay, but orbits above an altitude of about 500 miles or more expose astronauts to potentially harmful radiation associated with the van Allen belts. This can be mitigated by shielding, but that makes the spacecraft heavier. Orbits higher than about 5000 miles altitude stay above the van Allen belts, but take more fuel to reach. A typical circular low Earth orbit has an altitude of about 200 miles, where the figure makes clear that the problems of drag and radiation are minimized. An elliptical low Earth orbit might have a perigee of 100 miles and an apogee of 300 miles. Low Earth orbits are especially good for high resolution ground surveillance because they remain relatively close to the ground.

Visual satellite observers define LEO as being any orbit with a period of 225 minutes or less. Please see their excellent descriptions of orbits at their web site, www.satobs.org/faq/Chapter-04.txt.

Problem 13 The Period of Earth Orbits

What is the period of circular orbits with altitudes of 100, 200, and 300 miles? How much does it matter if you use statute or nautical miles?

Problem 14 The Altitude of a Satellite with a Period of 225 Minutes

As a brief review, calculate the altitude of a circular orbit that has a period of 225 minutes.

Geosynchronous Orbit or GEO

As circular orbits get larger their periods become longer, according to the formula from Chapter 1,

$$T_p = \frac{2\pi}{\sqrt{\mu}} a^{3/2}. \tag{3.1}$$

This formula applies to both elliptical and circular orbits, where, in the circular case, $a = r$. At some radius the period becomes equal to one day, and the satellite in such an orbit remains about one meridian on the Earth, provided that the orbit is direct. The plane of the satellite's orbit must include the center of the Earth, so the ground track of the satellite will oscillate about the equator (as in Figure 3.2-1 of the text) unless the orbital inclination is zero, in which case the ground track will, ideally, remain one point on the equator. This special case of a geosynchronous orbit is called geostationary. Because the plane of the orbit must include the center of the Earth the satellite cannot be made to hover above an arbitrary point on the Earth, but only above points on the equator.

Let's calculate the radius of such an orbit, which is

$$a^{3/2} = \frac{T_p \sqrt{\mu}}{2\pi}. \tag{3.2}$$

We want the period to be one sidereal day, or 86,164 seconds, rounded to the nearest whole second. In metric units, $\mu = 3.9860 \times 10^{14}$ m^3 s^{-2}, so

$a = 4.2164 \times 10^7$ meters. Subtracting the radius of the Earth, taken to be 6.3781×10^6 meters, leaves the altitude of the satellite as 3.5786×10^7 meters or 1.9323×10^4 nautical miles.

The importance of geostationary orbit for communications satellites is obvious, provided that the ground stations in use are not too far north or south. It may be useful for looking down at the Earth if high resolution is not required.

Mid Earth Orbit or MEO

Any satellite that is in an orbit between LEO and GEO must be in a mid Earth orbit, or MEO. Satellite tracking enthusiasts take MEO to be orbits with periods greater than or equal to 4 hours, but less than 24 hours. See the web page of the Canadian Satellite Tracking and Orbit Research organization at

www.castor2.ca/14_Orbits/02_MEO/index.html.

GPS satellite orbits and molniya satellites (more on these below) are in MEO.

High Earth Orbit or HEO

Satellites in orbits beyond GEO are said to be in HEO. The Moon is in HEO, as were the Apollo spacecraft sent to it. The Chandra X-Ray Observatory satellite is in HEO.

3.1.2 Classification of Orbits by Inclination

Polar Orbit

Orbits with inclinations of exactly 90 degrees are technically in polar orbit. Inclinations within a few degrees of 90 are also called polar. Polar orbits are particularly useful for satellites that must observe the entire surface of the Earth on a regular basis.

Low Inclination Orbit

This is not a technical term, but it implies inclinations near zero degrees.

3.1.3 Sun-Synchronous Orbits

We will see later in this chapter that the Earth's equatorial bulge causes orbits to evolve slowly with time. The specific motion causes the argument of perigee and the line of nodes to move. One can take advantage of this and launch a satellite into an orbit that takes the satellite over a given part of the Earth at the same Solar time every day, leading to the same solar illumination angle at every over-passage. This means that it is especially easy to compare series of images to look for small changes. Ground surveillance satellites are often launched into these orbits.

3.1.4 Special Orbits

Here are a few orbits that do not fit neatly into the above classifications.

Molniya Orbits

Molniya means lightning in Russian. It is used as the name of a type of orbit, a type of rocket, and a type of satellite. We consider the orbit.

Far northern countries, such as Russia and Canada, are not well served by communications satellites in geosynchronous orbits because they spend only half of their orbital period over the northern hemisphere. They are similarly poorly served by satellites in geostationary orbit because the satellites remain very far south whenever viewed from far northern locations.

These countries may be better served by satellites in highly eccentric, highly inclined orbits. The large eccentricity means that ground track of the satellite evolves slowly when the satellite is near apogee. If the apogee is far north then the northern country can make use of the satellite for a large fraction of the orbital period. This is the motivation for molniya orbits. Satellites in these orbits disappear from view very quickly and then return very quickly.

GPS Satellite Orbits

The reliability and success of the Global Positioning System requires that several satellites be accessible at all times from all locations on the Earth. This is accomplished by putting GPS satellites into circular orbits with periods of 12 hours. There must be at least two satellites in each orbit, with over-passage times equally separated. For reliability and redundancy more can be used.

Orbits Selected for Ground Station Passage

Satellites that collect a lot of data that must be relayed to a particular ground station or to a limited family of ground stations are often launched into an orbit specific to the needs of ground station over-passage.

3.2 The Earth's Equatorial Bulge

We saw in Chapter 2 that the Earth is an oblate spheroid, and that this makes it challenging to determine station coordinates. Here we will see that the oblate shape of the Earth causes non-central forces, leading to gravitational torque.

Conceptually, it is simplest to think of the equatorial bulge as a belt of additional mass around an otherwise spherical Earth. The spherical mass leads to central force and cannot cause gravitational torque. The belt can, and does, cause torque, as illustrated in Figure 3.1-4. The radius vectors and gravitational forces cause torque into the plane of the figure. Remember from your earlier classes that torque leads to a time rate of change of the angular momentum, $\vec{\tau} = \dot{\vec{L}}$. The figure then makes it clear that the angular momentum vector moves in a circle, causing regression of the line of nodes for direct orbits.

The text is a little weak on the details of what is going on, and in places the words do not match the figures, so I wish to fill in some details and quote results, but without the associated proofs. This is an advanced topic and we could easily spend half of the semester on it. The following results are adapted from "Introduction to Celestial Mechanics," by Jean Kovalevsky, copyright 1967, published by Springer-Verlag, New York. There are many books on celestial mechanics. I happen to have a copy of this one, given to me by Dr. Georgeanne Caughlan, one of my mentors in graduate school. I treasure it because she gave it to me. It may not be the best book on the subject, but it is my favorite for another reason.

The gravitational potential of the Earth is written as an expansion in Bessel functions, and the expansion is truncated to include only first-order terms. This allows solution for the time rate of change of Ω, the longitude of the ascending node, and ω, the argument of periapsis. This means that we have to think of the orbital elements as changing with time, which we will allow them to do. At any instant in the orbit the elements will have values of a_o, e_o, i_o, etc. In our notation the rates of change of Ω and ω are

$$\frac{d\Omega}{dt} = -\frac{3}{(1-e_o^2)^2}\frac{\mu^{1/2}}{a_o^{7/2}} J_2 \cos i_o, \tag{3.3}$$

and

$$\frac{d\omega}{dt} = \frac{1}{(1-e_o^2)^2}\frac{\mu^{1/2}}{a_o^{7/2}}\frac{3}{4} J_2 \left(5\cos^2 i_o - 1\right). \tag{3.4}$$

Most of the terms in these equations are familiar. One that may not be is J_2. This is a Bessel function of order two, and it looks like a decaying sine wave. Including its argument the function is $J_2(\frac{1}{2} - \frac{3}{2}\sin^2\phi)$, where ϕ is the angle between the current point in the orbit and the fundamental plane. Thus, ϕ tracks the object in its orbit, and J_2 varies smoothly as the object proceeds.

The leading negative sign in the equation for Ω shows that for direct orbits the line of nodes moves west. This is called regression of the line of nodes. The effect becomes weaker very quickly as a_o increases, which is not surprising, because the gravitational force of the equatorial bulge becomes more nearly central as a_o increases. I do find the dependence on inclination surprising, for on first look I would expect there to be no gravitational torque in the equatorial plane.

There is change in ω, called apsidial rotation, even if $e_o = 0$. The term in parentheses becomes zero when $\cos i_o = \sqrt{\frac{1}{5}}$, or when 63.4^0 and 116.6^0, explaining the result that is simply quoted in our text.

3.3 In-Plane Orbit Changes

3.3.1 Launching a Satellite and Adjusting its Orbit

A general launch of a satellite requires two events of burning the rocket engine, as in Figure 3.2-2 of the text. The first sends the satellite into an elliptical path that may be orbital, but is often suborbital, called the ascent ellipse. When the satellite reaches the apogee of this ellipse the rocket is fired again to send the satellite into the desired orbit. There can be small (or large!) errors in the launch azimuth, satellite speed, or flight-path angle that require adjustment of the orbit. This is done by making a small speed change called a Δv. General adjustment may require several Δvs.

Adjustment of Perigee and Apogee Height

In Chapter 1 we derived an energy equation that is valid for all conic-section orbits,

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}. \tag{3.5}$$

We solve this for v^2 to get

$$v^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right). \tag{3.6}$$

We use this to investigate the effect of changing v and its effect on a. Taking the differential,

$$2vdv = \frac{\mu}{a^2}da, \tag{3.7}$$

which gives

$$da = \frac{2a^2}{\mu}v\ dv. \tag{3.8}$$

For a very small change in velocity, dv, this equation allows us to calculate the change in the semi-major axis, da. The full major axis changes by twice this, or $2\ da$.

If we make the speed change at perigee then the result will be a change in the height of apogee. Similarly, if we apply the speed change at apogee then the result will be a change in perigee. Doing the math,

$$\Delta h_a \simeq \frac{4a^2}{\mu}v_p\ \Delta v_p, \tag{3.9}$$

and

$$\Delta h_p \simeq \frac{4a^2}{\mu}v_a\ \Delta v_a. \tag{3.10}$$

Maneuvers of this sort are often used for station keeping or maintaining a satellite in its desired orbit once it is already there. Station keeping usually requires relatively small corrections, but the use of fuel over time adds up, and has to be acknowledged in mission planning.

3.3.2 Hohmann Transfer

A Hohmann transfer orbit is an elliptical orbit that allows maneuvering a satellite from one coplanar circular orbit to another. It is a very useful and important maneuver. It is often the case that a satellite is launched into an initial, relatively low circular orbit to allow system checks in space, and then transferred to a higher, coplanar circular orbit. The Hohmann transfer is ideal for this. The actual transfer orbit is half of an ellipse that is tangent to each of the circular orbits. That the orbits are tangent is important, for this makes clear the direction in which the spacecraft must face when its rocket is fired. The Hohmann transfer orbit has the smallest Δv of all possible transfer orbits between the two circular orbits. It is named after Walter Hohmann, a German engineer who proposed it in a book published in 1925.

As in Figure 3.3-1 of the text, assume that we wish to transfer from a smaller circular orbit of radius r_1 to a larger, coplanar circular orbit of radius r_2. The transfer orbit is half of an ellipse and is clearly tangent to both circular orbits, so the rocket must be fired to speed up the spacecraft without changing its direction. From our earlier work we know the radius, major axis, and speed of each circular orbit, and we know the major axis of the elliptical orbit. Take the speed of the satellite in the smaller orbit to be v_{cs1} and that in the larger to be v_{cs2}. We also know from our earlier work that the satellite will have to speed up from v_{cs1} to enter the transfer ellipse, then speed up again at the far end of the ellipse to enter the larger circular orbit.

The major axis of the transfer ellipse is $2a_t = r_1 + r_2$. We can use this to calculate the specific energy of the transfer orbit as

$$\mathcal{E}_t = -\frac{\mu}{2a_t} = -\frac{\mu}{r_1 + r_2}. \tag{3.11}$$

We can now solve the energy equation

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} \tag{3.12}$$

for the velocity in the transfer ellipse when the radius is r_1 as

$$v_{t1} = \sqrt{2\left(\frac{\mu}{r_1} + \mathcal{E}_t\right)}. \tag{3.13}$$

Notice that I have departed slightly from the notation in the text to make clear that v_{t1} is the speed in the transfer orbit.

Just before firing the rocket to establish the transfer ellipse the satellite has the circular speed

$$v_{cs1} = \sqrt{\frac{\mu}{r_1}}. \tag{3.14}$$

Thus, the speed change required to establish the transfer ellipse is

$$\Delta v_1 = v_{t1} - v_{cs1}. \tag{3.15}$$

The satellite can coast from perigee to apogee, at which point the rocket must be fired again. The circular speed desired is

$$v_{cs2} = \sqrt{\frac{\mu}{r_2}}, \tag{3.16}$$

and the apogee speed of the transfer ellipse is

$$v_{t2} = \sqrt{2\left(\frac{\mu}{r_2} + \mathcal{E}_t\right)}. \tag{3.17}$$

Thus, the speed change required to enter the larger orbit is

$$\Delta v_2 = v_{cs2} - v_{t2}, \tag{3.18}$$

and the total is

$$\Delta v_{tot} = \Delta v_1 + \Delta v_2. \tag{3.19}$$

Problem 15 Hohmann Transfer from a Larger Circular Orbit to a Smaller One

Repeat the calculation above to transfer from the orbit with radius r_2 to the one with radius r_1. Note that this will require two speed decreases instead of increases. Does Δv_{tot} change? Why? What about the direction the satellite must face when its rocket is fired?

We should recognize that while the Hohmann transfer is efficient, it is also slow. The time of flight (TOF) is half of the orbital period of the whole ellipse, so

$$TOF = \pi\sqrt{\frac{a_t^3}{\mu}}. \tag{3.20}$$

Let's work an example.

Example 8 A communications satellite is in a circular orbit of radius 2 DU. Find the minimum Δv required to transfer to a circular orbit with double the initial altitude of the satellite.

The radius of the initial orbit is 2 DU, so the altitude is 1 DU. Doubling the altitude means moving to a circular orbit or radius 3 DU. Minimum Δv implies a Hohmann transfer.

The radius of the initial circular orbit is 2 DU, so

$$v_{cs1} = \sqrt{\frac{\mu}{r_1}} = \sqrt{\frac{1}{2}} = 0.7071 DUTU^{-1}. \tag{3.21}$$

For the transfer trajectory $r_p = 2$ DU and $r_a = 3$ DU, so $2a_t = 5$ DU, and

$$\mathcal{E}_t = -\frac{\mu}{2a_t} = -\frac{1}{5} DU^2 TU^{-2}, \tag{3.22}$$

which gives

$$v_{t1} = \sqrt{2\left(\frac{\mu}{r_1} + \mathcal{E}_t\right)} = \sqrt{\frac{3}{5}} = 0.7746 DUTU^{-1}. \tag{3.23}$$

This allows solution for the initial Δv,

$$\Delta v_1 = 0.7746 - 0.7071 = 0.0675 DUTU^{-1}. \tag{3.24}$$

When the satellite reaches apogee at the outer orbit of radius 3 DU its speed is

$$v_{t2} = \sqrt{2\left(\frac{\mu}{r_2} + \mathcal{E}_t\right)} = \sqrt{\frac{4}{15}} = 0.5163 DUTU^{-1}. \tag{3.25}$$

The speed of the circular orbit is

$$v_{cs2} = \sqrt{\frac{\mu}{r_2}} = \sqrt{\frac{1}{3}} = 0.5773 DUTU^{-1}, \tag{3.26}$$

and the second Δv is

$$\Delta v_2 = 0.5773 - 0.5163 = 0.0610 DUTU^{-1}, \tag{3.27}$$

which gives the total Δv as

$$\Delta v_{tot} = 0.0675 + 0.0610 = 0.1285 DUTU^{-1}. \tag{3.28}$$

The time of flight is

$$TOF = \pi\sqrt{\left(\frac{5}{2}\right)^3} = 12.42TU, \tag{3.29}$$

which seems like quite a long time. Compare this with half of the orbital period of each of the circular orbits,

$$\frac{T_{p1}}{2} = \pi\sqrt{2^3} = 8.89TU, \frac{T_{p2}}{2} = \pi\sqrt{3^3} = 16.32TU. \tag{3.30}$$

The Hohmann transfer orbit takes a long time because it has a long major axis.

Problem 16 A Generalization of Problem 3.8 in the Text

We need to work problems 3.1, 3.2, 3.3, 3.5, 3.6, 3.7, 3.8, and 3.9 in the text. Please do them in order. To get the maximum value from problem 3.8 we need to work it four times. The problem says, "Compute the minimum Δv required to transfer between two coplanar elliptical orbits which have their major axes aligned. The parameters for the ellipses are given by: $r_{p1} = 1.1DU, e_1 = 0.290, r_{p2} = 5DU, e_2 = 0.412$. Assume that both perigees lie on the same side of the Earth."

Let's work the initial problem twice, once using a transfer orbit that has its perigee at the apogee of the inner elliptical orbit and its apogee at the perigee of the outer orbit, and once using a transfer orbit that has its perigee at the perigee of the inner orbit and its apogee at the apogee of the outer orbit. Compare the results. Why are they different?

Now rotate the inner orbit by 180 degrees so that its apogee is on the opposite side of the Earth from the apogee of the outer orbit. Again consider two possible elliptical transfer orbits, one that has its perigee at the perigee of the inner orbit and its apogee at the perigee of the outer orbit, and another that has its perigee at the apogee of the inner orbit and its apogee at the apogee of the outer orbit. Compare the results for Δv.

3.3.3 General Coplanar Transfer

General coplanar transfers are faster than Hohmann transfers, but less efficient, and strict attention must be paid to the direction in which the rocket is fired. Attention must also be paid to the viability of a particular transfer orbit, which must intersect or be tangent to both the initial and final coplanar orbits. We will assume that the coplanar orbits are also circular.

Intersection of the transfer orbit with the circular orbits means that the periapsis of the transfer orbit must be equal to or smaller than the radius of the inner circular orbit, while at the same time the apoapsis of the transfer orbit must be equal to or larger than the radius of the outer circular orbit. These conditions are demonstrated in Figure 3.3-2 of the text, and are expressed mathematically as

$$r_p = \frac{p}{1+e} \leq r_1, \tag{3.31}$$

and

$$r_a = \frac{p}{1-e} \geq r_2, \tag{3.32}$$

where p and e are the parameter and eccentricity of the transfer orbit and r_1 and r_2 are the radii of the inner and outer circular orbits.

Figure 3.3-3 of the text shows a plot of the information in the two equations above with p on the horizontal axis and e on the vertical. This allows classification of possible transfer orbits by their eccentricity, which is a measure of shape, and indicates the family of possible transfer orbits. For the sake of investigation, or preliminary mission planning, assume that p and e for a viable hypothetical transfer orbit have been chosen for given circular orbits with radii r_1 and r_2. We know that the transfer orbit will work, but what are the directions in which the rocket must be fired, and what are the required Δvs?

Let's use what we know to calculate the specific energy and specific angular momentum of the transfer orbit. We know p and e, and want $\mathcal{E}_t$ and h_t. To find $\mathcal{E}_t$ we need a, which is $a = \frac{p}{1-e^2}$, so

$$\mathcal{E}_t = -\frac{\mu}{2a} = -\frac{\mu(1-e^2)}{2p}, \tag{3.33}$$

and to find h we remember that $p = h^2/\mu$, so

$$h_t = \sqrt{\mu p}. \tag{3.34}$$

We now follow the same procedure that we did in calculating the maneuvers in the Hohmann transfer. While in the smaller circular orbit the satellite already has speed

$$v_{cs1} = \sqrt{\frac{\mu}{r_1}}. \tag{3.35}$$

We solve the energy equation for the speed of the transfer orbit when its radius equals r_1 to get

$$v_{t1} = \sqrt{2\left(\frac{\mu}{r_1} + \mathcal{E}_t\right)}. \tag{3.36}$$

The angle between v_{t1} and v_{cs1} is the initial flight-path angle, ϕ_1. We can find this angle easily because, in general, $h = rv\cos\phi$, so

$$\cos\phi_1 = \frac{h_t}{r_1 v_{t1}}. \tag{3.37}$$

Figure 3.3-4 of the text shows the situation, where there is a vector triangle of sides v_{cs1}, v_{t1}, and Δv_1. We know the angle, ϕ_1, between the first two sides, so we can use the Law of Cosines for a plane triangle to calculate

$$\Delta v_1^2 = v_{t1}^2 + v_{cs1}^2 - 2v_{t1}v_{cs1}\cos\phi_1. \tag{3.38}$$

This is the first speed change and angle that we require.

The text implies that the second is left as an exercise for the reader. This is not acceptable to me because this is the first time that the topic has come up, so let's complete the calculation.

Following the pattern already started, the circular speed at radius r_2 is

$$v_{cs2} = \sqrt{\frac{\mu}{r_2}}, \tag{3.39}$$

and the speed of the transfer orbit at radius r_2 is

$$v_{t2} = \sqrt{2\left(\frac{\mu}{r_2} + \mathcal{E}_t\right)}. \tag{3.40}$$

The angle between v_{t2} and v_{cs2} is

$$\cos\phi_2 = \frac{h_t}{r_2 v_{t2}}, \tag{3.41}$$

so the required Δv to enter the circular orbit is

$$\Delta v_2^2 = v_{t2}^2 + v_{cs2}^2 - 2v_{t2}v_{cs2}\cos\phi_2. \tag{3.42}$$

The net is

$$\Delta v_{tot} = \Delta v_1 + \Delta v_2. \tag{3.43}$$

Let's try another example.

Example 9 Use the same initial and final circular orbits as in the section on Hohmann transfers - let the initial, lower orbit have a radius of 2 DU and the final, outer orbit have a radius of 3 DU, and compare the Δv in the general transfer orbit with that of the Hohmann transfer. In the earlier example we used a Hohmann transfer orbit with $2a = 5DU$. We calculated $\mathcal{E}_t$ for the transfer orbit. We could have calculated h_t, but didn't because it was not needed. Now to choose a new, general transfer orbit, it would be helpful to know p and e for the Hohmann transfer orbit. This is fairly simple, for the Hohmann transfer $r_p = 2DU = \frac{p}{1+e}$ and $r_a = 3DU = \frac{p}{1-e}$, and these serve as two equations in two unknowns for p and e,

$$p = 2(1+e) \tag{3.44}$$

and

$$p = 3(1-e). \tag{3.45}$$

These are easily solved to give $e = 0.2$ DU and $p = 2.4$ DU for the Hohmann transfer. There was also an equation that we derived back in Chapter 1,

$$e = \frac{r_a - r_p}{r_a + r_p} = \frac{3-2}{3+2} = \frac{1}{5} = 0.2, \tag{3.46}$$

which confirms the above result.

Now let's use this knowledge to choose a general transfer orbit and calculate its Δv. Figure 3.3-3 makes it clear that we could choose an ellipse, a parabola, or a hyperbola. No matter what the choice we must follow

$$r_p = \frac{p}{1+e} \le r_1 = 2DU, \tag{3.47}$$

and

$$r_a = \frac{p}{1-e} \ge r_2 = 3DU. \tag{3.48}$$

Let's choose an ellipse that is not too different from the Hohmann transfer ellipse, $r_p = 1.5DU, r_a = 3.5DU$. We easily find

$$2a = r_p + r_a = 5DU, \tag{3.49}$$

$$e = \frac{r_a - r_p}{r_a + r_p} = \frac{2}{5} = 0.4, \tag{3.50}$$

and

$$p = r_p(1+e) = 1.5 \times 1.4 = r_a(1-e) = 3.5 \times 0.6 = 2.1. \tag{3.51}$$

The specific energy of the transfer ellipse is

$$\mathcal{E}_t = -\frac{\mu}{2a} = -0.2DU^2TU^{-2}, \tag{3.52}$$

and the specific angular momentum is

$$h_t = \sqrt{\mu p} = \sqrt{1 \times 2.1} = 1.4491DU^2TU^{-1}. \tag{3.53}$$

The initial circular orbit has a velocity

$$v_{cs1} = \sqrt{\frac{\mu}{r_1}} = \sqrt{\frac{1}{2}} = 0.7071DUTU^{-1}, \tag{3.54}$$

and the transfer ellipse has a speed at radius $r_1 = 2DU$ of

$$v_{t1} = \sqrt{2\left(\frac{\mu}{r_1} + \mathcal{E}_t\right)} = \sqrt{0.6} = 0.7746DUTU^{-1}. \tag{3.55}$$

We note that the major axis length, $2a$ is the same for the Hohmann transfer ellipse and for the current transfer ellipse, so their energies are the same, so their speeds at radius $r_1 = 2DU$ are the same. Their values of Δv_1 are not the same, because the flight-path angle has changed. We calculate

$$\cos\phi_1 = \frac{h_t}{r_1 v_{t1}} = \frac{\sqrt{2.1}}{2 \times \sqrt{0.6}} = \sqrt{\frac{2.1}{2.4}} = 0.9354. \tag{3.56}$$

This means that ϕ_1 is about 20.7 degrees. Note that if we do the same calculation for the Hohmann transfer orbit at the same radius we get

$$\cos\phi_1 = \frac{h_t}{r_1 v_{t1}}, \tag{3.57}$$

but $h_t = r_1 v_{t1}$ because the radius and velocity are perpendicular, so $\cos\phi_1 = 1$ and $\phi_1 = 0$, as expected for an ellipse that is tangent to a circle.

We now calculate

$$\begin{aligned}\Delta v_1^2 &= v_{t1}^2 + v_{cs1}^2 - 2 v_{t1} v_{cs1} \cos\phi_1 \\ &= 0.6 + 0.5 - 2 \times 0.9354\sqrt{0.6 \times 0.5} = 0.07532 DU^2 TU^{-2}, \quad (3.58)\end{aligned}$$

which gives

$$\Delta v_1 = 0.2762 DUTU^{-1}. \quad (3.59)$$

We let the transfer orbit coast until the radius is equal to $r_2 = 3DU$. At that point the transfer speed is

$$v_{t2} = \sqrt{2\left(\frac{\mu}{r_2} + \mathcal{E}_t\right)} = \sqrt{\frac{4}{15}} = 0.5164 DUTU^{-1}, \quad (3.60)$$

and the circular speed is

$$v_{cs2} = \sqrt{\frac{\mu}{r_2}} = \sqrt{\frac{1}{3}} = 0.5774 DUTU^{-1}. \quad (3.61)$$

The angular momentum is conserved, so

$$\cos\phi_2 = \frac{h_t}{r_2 v_{t2}} = \frac{\sqrt{2.1}}{3\sqrt{\frac{4}{15}}} = \sqrt{\frac{2.1}{2.4}} = 0.9354, \quad (3.62)$$

so the flight-path angle is again about 20.7 degrees. This gives

$$\begin{aligned}\Delta v_2^2 &= v_{t2}^2 + v_{cs2}^2 - 2 v_{t2} v_{cs2} \cos\phi_2 \\ &= \frac{4}{15} + \frac{1}{3} - 2 \times 0.9354\sqrt{\frac{4}{15} \times \frac{1}{3}} = 0.04224 DU^2 TU^{-2}, \quad (3.63)\end{aligned}$$

which gives

$$\Delta v_2 = 0.2055 DUTU^{-1}. \quad (3.64)$$

The total is

$$\Delta v_{tot} = 0.4817 SDUTU^{-1}. \quad (3.65)$$

This is significantly larger than the Hohmann value of $0.1285 DUTU^{-1}$.

Before leaving this example let's consider a parabolic transfer orbit. The trajectory equation for a parabola is

$$r = \frac{p}{1+\cos\nu}, \tag{3.66}$$

where $e = 1$, so the minimum value of r is $p/2$, as we have seen before. Choose $p = 4$ DU, so $r_p = 2$ DU. This parabola is expected to be tangent to the inner orbit. Let's confirm this as part of the transfer orbit calculation. The inner circular orbit still has speed

$$v_{cs1} = \sqrt{\frac{\mu}{r_1}} = \sqrt{\frac{1}{2}} = 0.7071 DUTU^{-1}. \tag{3.67}$$

The parabola has zero total mechanical energy, so the speed in the parabola at any radius r is $v = \sqrt{2\mu/r}$, and at radius r_1

$$v_{t1} = \sqrt{\frac{2\mu}{r_{t1}}} = 1 DUTU^{-1}. \tag{3.68}$$

The transfer orbit has specific angular momentum

$$h_t = \sqrt{\mu p} = \sqrt{4} = 2DU^2TU^{-1}, \tag{3.69}$$

and we know that

$$\cos\phi_1 = \frac{h_t}{r_1 v_{t1}} = \frac{2}{2\times 1} = 1, \tag{3.70}$$

so $\phi_1 = 0$ as expected, and the parabolic orbit is tangent to the circular orbit. We can now calculate

$$\Delta v_1 = 1 - 0.7071 = 0.2929 DUTU^{-1}. \tag{3.71}$$

We note that this is about 10 percent larger than Δv_1 for the elliptical transfer example just worked.

Firing the rocket establishes the parabolic transfer orbit, which is allowed to proceed until the radius is equal to r_2. At that point

$$v_{t2} = \sqrt{\frac{2\mu}{r_{t2}}} = \sqrt{\frac{2}{5}} = 0.6324 DUTU^{-1}, \tag{3.72}$$

where the local circular speed is

$$v_{cs2} = \sqrt{\frac{\mu}{r_2}} = \sqrt{\frac{1}{5}} = 0.4472 DUTU^{-1}. \tag{3.73}$$

This time there is no reason to think that the orbits are tangent, so we must establish the angle and use it to find Δv_2. The specific angular momentum is constant, so

$$\cos\phi_2 = \frac{h_t}{r_2 v_{t2}} = \frac{2}{5 \times 0.6324} = 0.6325, \tag{3.74}$$

so ϕ_2 is about 50.7 degrees. This is a rather large angle, making the cosine term rather small, and because the cosine term in the formula for Δv is negative it means that the maneuver will be expensive of fuel. Using the law of cosines

$$\begin{aligned}\Delta v_2^2 &= v_{t2}^2 + v_{cs2}^2 - 2 v_{t2} v_{cs2} \cos\phi_2 \\ &= \frac{2}{5} + \frac{1}{5} - 2\sqrt{\frac{2}{25}} \cos\phi_2 = 0.8578 DU^2 TU^{-2}, \end{aligned} \tag{3.75}$$

so $\Delta v_2 = 0.9262 DUTU^{-1}$, and

$$\Delta v_{tot} = \Delta v_1 + \Delta v_2 = 0.2929 + 0.9262 = 1.2191 DUTU^{-1}. \tag{3.76}$$

This is the most efficient parabolic transfer orbit because it is tangent to the circle, but it is very expensive of fuel compared with the Hohmann transfer.

Time of Flight

The parabolic transfer orbit might be advantageous if it takes significantly less time than the Hohmann transfer. We don't yet know how to calculate the time of flight for two arbitrary points in any conic-section trajectory. That will be our goal in Chapter 4.

3.4 Bi-elliptic Transfer

This subject is not covered in the text, but it is included in homework problems 9 and 10. Let's cover enough so that the homework makes sense.

Bi-elliptic transfer is interesting because sometimes it requires a smaller Δv than a Hohmann transfer, but it does this at the expense of taking a great deal more time. Consider the case of transfer from a small circular orbit of radius r_1 to a larger circular radius of radius r_2. We can work the

Δv for the Hohmann transfer in closed form, and we will need the result for comparison with bi-elliptic transfer. The circular speed of the inner orbit is

$$v_{cs1} = \sqrt{\frac{1}{r_1}}. \tag{3.77}$$

Similarly, for the outer circular orbit

$$v_{cs2} = \sqrt{\frac{1}{r_2}}. \tag{3.78}$$

The transfer ellipse has

$$2a_t = r_1 + r_2, \tag{3.79}$$

giving

$$\mathcal{E}_t = -\frac{1}{r_1 + r_2}. \tag{3.80}$$

This allows calculation of the velocity of the transfer orbit at radii r_1 and r_2, respectively, as

$$v_{t1} = \sqrt{2\left(-\frac{1}{r_1 + r_2} + \frac{1}{r_1}\right)} = \sqrt{\frac{2r_2}{r_1(r_1 + r_2)}}, \tag{3.81}$$

and

$$v_{t2} = \sqrt{2\left(-\frac{1}{r_1 + r_2} + \frac{1}{r_2}\right)} = \sqrt{\frac{2r_1}{r_2(r_1 + r_2)}}. \tag{3.82}$$

These results make sense because we can convert from v_{t1} to v_{t2} by simply interchanging the subscripts 1 and 2. From them we find that the rocket engine must be fired to speed up from the initial circular orbit to the transfer orbit by an amount

$$\Delta v_1 = \sqrt{\frac{2r_2}{r_1(r_1 + r_2)}} - \sqrt{\frac{1}{r_1}}. \tag{3.83}$$

Similarly, the rocket engine must be fired to speed up from the transfer ellipse to the outer orbit by an amount

$$\Delta v_2 = \sqrt{\frac{1}{r_2}} - \sqrt{\frac{2r_1}{r_2(r_1 + r_2)}}. \tag{3.84}$$

This gives Δv_{tot} of

$$\Delta v_{tot} = \sqrt{\frac{2r_2}{r_1(r_1+r_2)}} - \sqrt{\frac{1}{r_1}} + \sqrt{\frac{1}{r_2}} - \sqrt{\frac{2r_1}{r_2(r_1+r_2)}}. \tag{3.85}$$

We also have

$$TOF = \pi\sqrt{\left(\frac{r_1+r_2}{2}\right)^3}. \tag{3.86}$$

Plugging in the numbers from our recent example, $r_1 = 2DU$, $r_2 = 3DU$, gives

$$\Delta v_{tot} = \sqrt{\frac{3}{5}} - \sqrt{\frac{1}{2}} + \sqrt{\frac{1}{3}} - \sqrt{\frac{4}{15}} = 0.1285 DUTU^{-1}, \tag{3.87}$$

all of which looks very familiar, and

$$TOF = \pi\sqrt{\left(\frac{5}{2}\right)^3} = 12.42TU. \tag{3.88}$$

Homework problem 9 in the text leads us to think that bi-elliptic transfer is transfer to a very large circular orbit, but this is inconsistent with the name and is not the simplest way to think about what happens. Instead, think of bi-elliptic transfer as first burning the rocket engine to transfer from a circular orbit of radius r_1 to a very large elliptical orbit with apocenter at r_b and pericenter at r_1 by speeding up. Upon reaching apocenter the rocket engine is burned again to transfer to a second elliptical orbit (hence the name bi-elliptical) with apocenter at that same r_b and pericenter at r_2 by speeding up again. Upon reaching r_2 the rocket is fired a third time to transfer to a circular orbit of radius r_2 by slowing down.

The circular orbits have speeds

$$v_{cs1} = \sqrt{\frac{1}{r_1}}, \tag{3.89}$$

and

$$v_{cs2} = \sqrt{\frac{1}{r_2}}. \tag{3.90}$$

The first transfer ellipse has specific mechanical energy

$$\mathcal{E}_{t1} = -\frac{1}{r_1+r_b}, \tag{3.91}$$

giving it pericenter and apocenter velocities

$$v_{t11} = \sqrt{\frac{2r_b}{r_1(r_1 + r_b)}}, \tag{3.92}$$

and

$$v_{t1b} = \sqrt{\frac{2r_1}{r_b(r_1 + r_b)}}, \tag{3.93}$$

respectively. Note the convention being built for the subscripts. The subscript t indicates transfer, then the following numeral, 1 or 2, denotes transfer ellipse 1 or 2, and the final subscript denotes the radius where the velocity applies, 1, 2, or b. The second transfer ellipse has energy

$$\mathcal{E}_{t2} = -\frac{1}{r_2 + r_b}, \tag{3.94}$$

giving it pericenter and apocenter velocities

$$v_{t2b} = \sqrt{\frac{2r_2}{r_b(r_2 + r_b)}}, \tag{3.95}$$

and

$$v_{t22} = \sqrt{\frac{2r_b}{r_b(r_2 + r_b)}}, \tag{3.96}$$

respectively.

To transfer from the inner circle to the first ellipse at radius r_1

$$\Delta v_1 = \sqrt{\frac{2r_b}{r_1(r_1 + r_b)}} - \sqrt{\frac{1}{r_1}}, \tag{3.97}$$

then to transfer from the first ellipse to the second ellipse at radius r_b

$$\Delta v_b = \sqrt{\frac{2r_2}{r_b(r_2 + r_b)}} - \sqrt{\frac{2r_1}{r_b(r_1 + r_b)}}, \tag{3.98}$$

and to transfer from the second ellipse to the outer circle at radius r_2

$$\Delta v_2 = \sqrt{\frac{2r_b}{r_2(r_2 + r_b)}} - \sqrt{\frac{1}{r_2}}. \tag{3.99}$$

In calculating Δv_2 care was taken to recognize that the speed of the outer circular orbit is slower than that of the second ellipse.

There is a worked example concerning Earth orbit on pages 329 and 330 of Fundamentals of Astrodynamics and Applications by Vallado (2007). We take $r_1 = 1.03DU$, $r_b = 80DU$, and $r_2 = 60DU$. We find

$$\Delta v_1 = 1.384581 - 0.985329 = 0.399253DUTU^{-1}, \tag{3.100}$$

$$\Delta v_b = 0.103510 - 0.017826 = 0.085683DUTU^{-1}, \tag{3.101}$$

and

$$\Delta v_2 = 0.138013 - 0.129099 = 0.008136DUTU^{-1}. \tag{3.102}$$

The net is

$$\Delta v = 0.493DUTU^{-1} \tag{3.103}$$

in agreement with Vallado. The time for this transfer is

$$TOF = \pi\sqrt{40.515^3} + \pi\sqrt{70^3} = 2650.076TU, \tag{3.104}$$

again in agreement with Vallado.

We compare these results with the Hohmann transfer, and now it is very handy to have the Hohmann results in closed form. We get

$$\begin{aligned}\Delta v &= \sqrt{\frac{120}{1.03(61.03)}} - \sqrt{\frac{1}{1.03}} + \sqrt{\frac{1}{60}} - \sqrt{\frac{2.06}{60(61.03)}} \\ &= 1.381657 - 0.985329 + 0.129099 - 0.0237184 = 0.501709DUTU^{-1},\end{aligned} \tag{3.105}$$

which is a little larger than the bi-elliptic result, but

$$TOF = \pi\sqrt{\left(\frac{61.03}{2}\right)^3} = 529.566TU, \tag{3.106}$$

which is much shorter than the bi-elliptic transfer time.

We note that as r_b approaches infinity the velocity changes become

$$\Delta v_1 = \sqrt{\frac{2}{r_1}} - \sqrt{\frac{1}{r_1}}, \tag{3.107}$$

$$\Delta v_b = 0, \tag{3.108}$$

and

$$\Delta v_2 = \sqrt{\frac{2}{r_2}} - \sqrt{\frac{1}{r_2}} \qquad (3.109)$$

or

$$\Delta v = \left(\sqrt{2} - 1\right)\left(\sqrt{\frac{1}{r_1}} + \sqrt{\frac{1}{r_2}}\right). \qquad (3.110)$$

3.5 Out-Of-Plane Orbit Changes

So far we have considered only the magnitude of Δv, and we have set up a situation where $\Delta\vec{v}$ is only along the direction of motion and in the orbital plane. Such maneuvers can change the orbit's size or shape and rotate the line of apsides. We now explicitly consider the vector nature of $\Delta\vec{v}$ to change the plane of the orbit, that is, to change the orbital inclination.

The change of plane of a satellite's orbit is most easily analyzed and accomplished when the satellite is on the line of nodes, at either the ascending or descending node. Consider the situation in Figure 3.4-1 in the text, where we analyze how to change an inclined orbit into an equatorial one. The maneuver is made when the satellite is at or near the descending node in this case. A similar maneuver with an oppositely directed $\Delta\vec{v}$ is possible at the ascending node. The figure shows an isosceles triangle, so only the plane of the orbit changes, not the other parameters. The triangle of velocities is easily analyzed to give the magnitude of Δv,

$$\Delta v = 2v \sin\frac{\theta}{2}, \qquad (3.111)$$

where θ is the angle through which the orbital inclination moves. This equation shows that large changes in θ are expensive of fuel, even more so if the orbit is low and fast. This is the problem faced by far northern countries. Once again, we will investigate the details of time of flight between general points in elliptical orbits in Chapter 4, which will enable us to calculate the properties of molniya orbits.

The equation just derived is not specific about the direction of Δv, so let's find the direction. As mentioned before, the triangle in Figure 3.4-1 is isosceles. We consider only one half of the triangle, so that we have a right triangle with two known angles, $\frac{\theta}{2}$ at the skinny apex and $\frac{\pi}{2}$ radians or 90 degrees at the base opposite. The remaining angle must be $\frac{\pi}{2} - \frac{\theta}{2} = \frac{\pi-\theta}{2}$ in radians, or $\frac{180-\theta}{2}$ in degrees. This is the angle of $\Delta\vec{v}$ relative to the velocity vector $\vec{v}$ in the initial, inclined orbit.

Chapter 4

$\vec{r}$ and $\vec{v}$ as Functions of Time

Back in Chapter 1 we found the equation of motion for the equivalent one-body problem of the two-body, gravitational motion problem to be

$$\ddot{\vec{r}} = -\frac{\mu}{r^3}\vec{r}. \tag{4.1}$$

Usually, upon writing a second-order differential equation like this, we expect to find a solution of the form $\vec{r}(t)$, or $(x(t), y(t), z(t))$, or $(r(t), \theta(t), \phi(t))$, and to find $\vec{v}(t)$ along the way. For reasonably behaved functions we expect that $r(t)$ and $v(t)$ will be invertible, so that we can solve for $t(r)$ and $t(v)$. We have not done this yet. This is the goal of Chapter 4.

4.1 What we Have Done

Instead, of finding $\vec{r}(t)$ and $\vec{v}(t)$, we have learned all that we can about them without actually finding them. We have found that $\mathcal{E}, \vec{h}$, and $\vec{B}$ are constants of the motion. From this we have found that $\mathcal{E}$ determines the orbit's semi-major axis, a, from

$$\mathcal{E} = -\frac{\mu}{2a}, \tag{4.2}$$

and that $\mathcal{E}$ determines the magnitude of the velocity, $v(r)$, from

$$v = \sqrt{2\left(\mathcal{E} - \frac{\mu}{r}\right)}. \tag{4.3}$$

We have found that the conservation of $\vec{h}$ determines the plane of the orbit, so we need consider only two coordinates, r and ν, and that $\vec{h}$ also determines the semi-latus rectum or parameter of the orbit from

$$p = \frac{h^2}{\mu}. \tag{4.4}$$

We have found that $\vec{B}$ determines the eccentricity or shape of the orbit and the direction to the pericenter from

$$\vec{e} = \frac{\vec{B}}{\mu} = \frac{1}{\mu}\left[\left(v^2 - \frac{\mu}{r}\right)\vec{r} - \left(\vec{r}\cdot\vec{v}\right)\vec{v}\right]. \tag{4.5}$$

Finally, we have found that multiplying the equation of motion by $\vec{h}$ in a vector cross product and simplifying gives the the relationship between r and ν as the trajectory equation,

$$r = \frac{p}{1 + e\cos\nu}. \tag{4.6}$$

4.2 Elliptical Time of Flight as a Function of E

Be prepared for a new vocabulary for describing orbital motion - new in the sense of new to you. The vocabulary that we are about to learn is very much the vocabulary of Johannes Kepler. Kepler's concepts were developed to describe only elliptical motion, largely because he was unaware of parabolic or hyperbolic motion, but the concepts are easily generalized to include them.

We already know that the true anomaly is the angle measured in the direction of motion from pericenter to the location of the satellite at the desired time, called the epoch. The occupied focus is used as the apex for measuring this angle. Kepler found that the equation for time of flight in an orbit could be simplified by defining the eccentric anomaly, E. This uses the orbit's center as the location of the measured angle, and measures from pericenter to a location on an auxiliary circle circumscribing the ellipse.

4.2.1 Two Approaches: Geometric and Analytical

Kepler's Second Law provides the initial basis for bringing time into our work: "The line joining any planet to the Sun sweeps out equal areas in equal times." This applies to all conics, not just ellipses, because the force in the idealized two-body problem is central and cannot cause torques. Kepler used this to develop an equation for time of flight in an elliptical orbit based on geometry.

We start by drawing an elliptical orbit, marking its center and occupied focus, and circumscribing a circle. We already know the true anomaly, or angle ν, from the pericenter to an arbitrary point on the orbit. Kepler's Second Law says that the rate at which the line from the focus to the orbiting body sweeps out area is constant. We can write this mathematically as

$$\frac{t-T}{T_p} = \frac{\mathcal{A}_1}{\pi ab}, \tag{4.7}$$

where t is the time at which the orbiting body is at true anomaly ν, T is the time at which the orbiting body is at true anomaly $\nu = 0$, or pericenter, T_p is the period of the elliptical orbit, $\mathcal{A}_1$ is the area swept out by the radius vector as the orbiting body moves from $\nu = 0$ to position ν, and πab is the area of the ellipse. We can solve for the time of flight alone to get

$$t - T = \frac{T_p}{\pi ab}\mathcal{A}_1. \tag{4.8}$$

Time of flight can also be defined directly from the conservation of angular momentum using differential calculus, which was unknown to Kepler. We write

$$h = rv_t, \tag{4.9}$$

where v_t is the transverse component of the velocity, which is needed to calculate the magnitude of a cross product. If we think of the area swept out by the radius vector during a small interval of time then we can write

$$v_t = r\dot{\nu} = r\frac{d\nu}{dt}, \tag{4.10}$$

so

$$h = r^2\frac{d\nu}{dt}. \tag{4.11}$$

Working with differentials, this gives

$$hdt = r^2 d\nu. \tag{4.12}$$

We can integrate both sides from pericenter to the true anomaly ν at time t to give

$$\int_T^t h dt = \int_0^\nu r^2 dv, \tag{4.13}$$

where we have again labeled the time of pericenter passage at T. We know that h is a constant, so it may be factored out of the integral on the left-hand side, making that integral simple to evaluate. We can replace r in the integral on the right hand side with its value from the trajectory equation to get

$$h(t-T) = \int_0^\nu \frac{p^2}{(1+e\cos\nu)^2} d\nu. \tag{4.14}$$

This integral applies to all conic-section orbits. We will see that it can be very challenging to evaluate.

The two approaches, one based on Kepler's Second Law and the other on conservation of momentum (which is the reason why Kepler's Second Law holds), must give the same results for $t-T$. We will spend some time showing this and generalizing the results for parabolic and hyperbolic orbits.

4.2.2 Why is This Integral so Difficult?

The integral for time of flight is difficult for two reasons. First, the integral for the arc length of circumference of an ellipse is difficult. Consider a differential length of arc on an ellipse. We have

$$d\ell = \sqrt{dx^2 + dy^2}, \tag{4.15}$$

where $d\ell$ is a differential bit of arc length. This is familiar and simple enough. The equation for the ellipse in Cartesian coordinates is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{4.16}$$

Solving for y gives

$$y = \pm b\sqrt{1 - \frac{x^2}{a^2}}. \tag{4.17}$$

The two possible signs designate the parts of the ellipse above and below the x axis. At this point we can recognize that the symmetry of the ellipse gives four arcs of equal length, and that we need to find the length of only one to know the circumference of the ellipse. Thus, we can choose the positive sign and proceed. Taking the differential gives

$$dy = \frac{-bx}{a^2} \frac{1}{\sqrt{1 - \frac{x^2}{a^2}}} dx. \tag{4.18}$$

Squaring, substituting, reorganizing, and integrating gives a method to find the arc length

$$\ell = \int d\ell = \int_0^a \sqrt{1 + \frac{b^2x^2}{a^2}\left(\frac{1}{1 - \frac{x^2}{a^2}}\right)} dx. \tag{4.19}$$

This integral is not easy, and it was not evaluated until Euler found a series solution in 1732 and published in 1738. See the excellent web site by Sandifer at
www.maa.org/editorial/euler/How%20Euler%20Did%20It%2012%20arc%20length%20ellipse.pdf

The second reason that it is difficult to do the time of flight integral is that the orbiting body is constantly changing speed. Thus, a technique based on the formula

$$dt = \frac{d\ell}{v} \tag{4.20}$$

is difficult because both $d\ell$ and v contribute to the difficulty.

4.2.3 Kepler's Geometric Method - Developing Kepler's Equation

Consider Figure 4.2-2 of the text, which shows the elliptical orbit, the circumscribed circle, the true anomaly, the eccentric anomaly, the area $\mathcal{A}_1$, and several additional helpful points and areas. In passing we note that the eccentric anomaly seems like a misnomer. Given that the apex is at the center of the circle we might expect that E would be called the centric anomaly, but it is not.

We already know that for a fixed value of x the ratio of the height of the ellipse to that of the circle is $\frac{b}{a}$. This will be helpful in finding an expression for the area $\mathcal{A}_1$.

First we recognize that

$$\mathcal{A}_1 = \mathcal{A}_{PSV} - \mathcal{A}_2. \tag{4.21}$$

The area of triangle $\mathcal{A}_2$ is relatively easy to find, because the triangle's height is $\frac{b}{a}a\sin E$ and its base is $ae - a\cos E$, so

$$\mathcal{A}_2 = \frac{ab}{2}\big(e\sin E - \cos E\sin E\big). \tag{4.22}$$

Area $\mathcal{A}_{PSV}$ is just $\frac{b}{a}$ times area $\mathcal{A}_{QSV}$. Area $\mathcal{A}_{QSV}$ is the area of the sector of the circle QOV minus the area of the triangle QOS. The area of the sector of a circle is easily developed into angular measure of area. The area of sector QOV divided by the area of the entire circle must equal angle e divided by 2π. Thus

$$\frac{\mathcal{A}_{QOV}}{\pi a^2} = \frac{E}{2\pi}, \tag{4.23}$$

or

$$\mathcal{A}_{QOV} = \frac{a^2}{2}E. \tag{4.24}$$

Triangle QOS has a base of $a\cos E$ and a height of $a\sin E$, so its area is

$$\mathcal{A}_{QOS} = \frac{a^2}{2}\cos E\sin E, \tag{4.25}$$

so

$$\begin{aligned}\mathcal{A}_{PSV} = \frac{b}{a}\Bigg(\mathcal{A}_{QOV} - \mathcal{A}_{QOS}\Bigg) \\ = \frac{b}{a}\Bigg(\frac{a^2}{2}E - \frac{a^2}{2}\cos E sin E\Bigg) \\ = \frac{ab}{2}\Bigg(E - \cos E\sin E\Bigg).\end{aligned} \tag{4.26}$$

We now calculate

$$\mathcal{A}_1 = \frac{ab}{2}\Bigg(E - \cos E\sin E - e\sin E + \cos E\sin E\Bigg) = \frac{ab}{2}\Bigg(E - e\sin E\Bigg). \tag{4.27}$$

Recognizing that $T_p = 2\pi\sqrt{\frac{a^3}{\mu}}$ gives the time of flight as

$$t - T = \frac{T_p}{\pi ab}\mathcal{A}_1 = \sqrt{\frac{a^3}{\mu}}\Bigg(E - e\sin E\Bigg). \tag{4.28}$$

This is often written using Kepler's notation

$$M = E - e \sin E, \tag{4.29}$$

with the mean motion defined as

$$n = \sqrt{\frac{\mu}{a^3}} \tag{4.30}$$

so that

$$M = n(t - T) = E - e \sin E, \tag{4.31}$$

which is known as Kepler's equation. It is called transcendental or transcendental of algebra because it cannot be solved algebraically. Graphical and numerical methods must be used.

Note that because we now have the function $t(E)$ that we can, presumably, invert to find $E(t)$. We still don't have $E(t)$ explicitly, nor do we have $\nu(t)$ nor $r(t)$.

Relating ν and E

To make Kepler's equation useful we must be able to relate the true anomaly to the eccentric anomaly, so we need $E(\nu)$ and $\nu(E)$. Going back to Figure 4.2-2, it is relatively easy to derive the cosine of E using triangle QOS that we have already analyzed. Its base is $ae + r\cos\nu$ and its hypotenuse is a, so

$$\cos E = \frac{ae + r\cos\nu}{a} = e + \frac{r}{a}\cos\nu. \tag{4.32}$$

We can replace r with its value from the trajectory equation

$$r = \frac{a(1 - e^2)}{1 + e\cos\nu} \tag{4.33}$$

or

$$\frac{r}{a} = \frac{1 - e^2}{1 + e\cos\nu} \tag{4.34}$$

to get

$$\cos E = \frac{e(1 + e\cos\nu)}{1 + e\cos\nu} + \frac{(1 - e^2)\cos\nu}{1 + e\cos\nu} = \frac{e + \cos\nu}{1 + e\cos\nu}. \tag{4.35}$$

This allows calculation of E from ν. We can solve this result for $\cos\nu$ to calculate ν from E.

$$\cos E(1 + e\cos\nu) = e + \cos\nu, \tag{4.36}$$

so

$$\cos E + e\cos E\cos\nu = e + \cos\nu, \tag{4.37}$$

and

$$\cos E - e = \cos\nu(1 - e\cos E), \tag{4.38}$$

with the desired result

$$\cos\nu = \frac{\cos E - e}{1 - e\cos E}. \tag{4.39}$$

The equations for $E(\nu)$ and $\nu(E)$ can be rewritten in another form that makes them potentially more useful. The following derivation comes from the web site for the course AA 4362 in Astrodynamics at the Naval Postgraduate School taught by Dr. Stephen A. Whitmore in 2002. The web site is web.nps.navy.mil/s̃sweb/AA4362/AA4362.html

Under the guiding philosophy that "equals done to equals gives equals", given the equation above for $\cos E$ as a function of $\cos\nu$, we can write

$$\frac{1-\cos E}{1+\cos E} = \frac{1 - \frac{\cos\nu + e}{1+e\cos\nu}}{1 + \frac{\cos\nu + e}{1+e\cos\nu}} = \frac{1 + e\cos\nu - \cos\nu - e}{1 + e\cos\nu + \cos\nu + e} = \frac{1-e}{1+e}\frac{1-\cos\nu}{1+\cos\nu}. \tag{4.40}$$

For any angle α there are trig identities

$$1 - \cos\alpha = 2\sin^2\frac{\alpha}{2}$$

$$1 + \cos\alpha = 2\cos^2\frac{\alpha}{2}. \tag{4.41}$$

Applying this to the equation above gives

$$\tan^2\frac{E}{2} = \frac{1-e}{1+e}\tan^2\frac{\nu}{2}. \tag{4.42}$$

Taking the square root gives the desired pair of results

$$\tan\frac{E}{2} = \sqrt{\frac{1-e}{1+e}}\tan\frac{\nu}{2},$$

$$\tan\frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}. \tag{4.43}$$

We now have the needed suite of relations between E and ν. Note that they are all transcendental of algebra. We note from the drawing of the elliptical orbit and its circumscribing circle that E and ν are always in the same half plane, so when $0 \leq \nu \leq 180^o$ so is E.

For good measure, and in anticipation of future need, let's derive $\sin\nu$ as a function of $\sin E$. Referring to Figure 4.2-2 in the text again, we see that the sine of ν is the distance PS divided by r, and PS is the fraction $\frac{b}{a}$ times the distance QS, which is $a\sin E$, so

$$\sin\nu = \frac{b}{r}\sin E. \tag{4.44}$$

Remember from Chapter 1 that

$$b = a\sqrt{1-e^2}, \tag{4.45}$$

so

$$\sin\nu = \frac{a\sqrt{1-e^2}}{r}\sin E. \tag{4.46}$$

4.2.4 Time of Flight Between Arbitrary Points

We wish to be able to find the time of flight between an initial point in the orbit at time t_o and true anomaly ν_o that is not necessarily at pericenter and some general point ν. If the orbiting body does not pass through pericenter then we can do this with the formalism that we already have because

$$t - t_o = (t - T) - (t_o - T). \tag{4.47}$$

If the orbiting body does pass through pericenter, because $\nu_o > \nu$, then

$$t - t_o = T_p + (t - T) - (t_o - T). \tag{4.48}$$

Cast into the form using the eccentric anomaly gives

$$t - t_o = \sqrt{\frac{a^3}{\mu}}\Big(2k\pi + (E - e\sin E) - (E_o - e\sin E_o)\Big), \tag{4.49}$$

where k is the integer number of times that the orbiting body passes through pericenter.

Let's try an example.

Example 10 An Elliptical Orbit

A space probe is in an elliptical orbit about the Sun. Perihelion distance is 0.5 AU and aphelion distance is 2.5 AU. How many days in each orbit is the probe closer than 1.0 AU to the Sun?

Clearly we need to use a version of Kepler's equation, so we need to put the information that we have into the from that we need. The probe does not make a complete orbit about the Sun in the problem, so we need to apply

$$t - T = \sqrt{\frac{a^3}{\mu}}\Big((E - e\sin E) - (E_o - e\sin E_o)\Big) \tag{4.50}$$

for some appropriate choice of E and E_o. The orbit is symmetric about perihelion, so let's choose E_o at perihelion, E at 1 AU, caclulate $t-T$ between those points, and double the result to answer the question.

First, let's find the eccentricity, e, from the data given. Way back in Chapter 1 we derived an equation that is valuable now

$$e = \frac{r_a - r_p}{r_a + r_p} = \frac{2}{3}. \tag{4.51}$$

Similarly, we can find the semi-major axis from

$$a = \frac{r_a + r_p}{2} = \frac{3}{2}. \tag{4.52}$$

A form of the trajectory equation,

$$r = \frac{a(1-e^2)}{1 + e\cos\nu} \tag{4.53}$$

can be solved for ν,

$$\cos\nu = \frac{a(1-e^2) - r}{er}. \tag{4.54}$$

At perihelion this is especially easy to apply, for $r = r_a = 0.5$ AU, and we know that ν_o should be zero. Let's confirm this as a way to develop some experience,

$$\cos\nu_o = \frac{\frac{3}{2}\left(1 - \frac{4}{9}\right) - \frac{1}{2}}{\frac{2}{3}\frac{1}{2}} = \frac{\frac{15}{18} - \frac{1}{2}}{\frac{1}{3}} = 1, \tag{4.55}$$

which gives $\nu_o = 0$ as expected. When ν_o is zero E_o is also zero. When $r = 1.0$ AU

$$\cos\nu = \frac{\frac{3}{2}\left(1 - \frac{4}{9}\right) - 1.0}{\frac{2}{3}} = \frac{-\frac{3}{18}}{\frac{2}{3}} = -\frac{1}{4}. \tag{4.56}$$

We then use

$$\cos E = \frac{e + \cos\nu}{1 + e\cos\nu} = \frac{\frac{2}{3} - \frac{1}{4}}{1 - \frac{2}{3}\frac{1}{4}} = \frac{\frac{8}{12} - \frac{3}{12}}{\frac{12}{12} - \frac{2}{12}} = \frac{1}{2}. \tag{4.57}$$

We know that this occurs at an eccentric anomaly between zero and ninety degrees, so $E = 60^o = \frac{\pi}{3}$ radians, and

$$\sin E = \frac{\sqrt{3}}{2} = 0.866. \tag{4.58}$$

From this we find the time from perihelion to 1.0 AU as

$$t - T = \sqrt{\frac{1.5^3}{\mu}}\left(\frac{\pi}{3} - \frac{2}{3}\frac{\sqrt{3}}{2}\right) = 0.863TU. \tag{4.59}$$

The time that we want is double this,

$$t_{tot} = TOF = 1.726TU. \tag{4.60}$$

This turns out to be 1.726 TU times 58.133 days per TU, or 100.36 days. For comparison, the time for a complete orbit is

$$T_p = 2\pi\sqrt{\frac{a^3}{\mu}} = 11.543TU = 671.03days. \tag{4.61}$$

4.2.5 Analytical Method

We return to the equation

$$\int_T^t hdt = h(t - T) = \int_0^\nu r^2 d\nu. \tag{4.62}$$

We view the transformation from ν to E as a simple change of variable. First we recognize that the lower limit of integration remains zero because $E = 0$ when $\nu = 0$, and the upper limit transforms from ν to E. We will use the relation between E and ν to find a general relation between r and E, and differentiate it to find a relation between $d\nu$ to dE. We recall that

$$\cos\nu = \frac{\cos E - e}{1 - e\cos E}, \tag{4.63}$$

so

$$\begin{aligned}
r = \frac{p}{1+e\cos\nu} &= \frac{p}{1+\frac{e\cos E-e^2}{1-e\cos E}} \\
&= \frac{p}{\frac{1-e\cos E+e\cos E-e^2}{1-e\cos E}} \\
&= \frac{p(1-e\cos E)}{1-e^2} = a(1-e\cos E). \qquad (4.64)
\end{aligned}$$

Differentiating the equation for $\cos\nu$ gives

$$-\sin\nu d\nu = -\frac{\sin E dE}{1-e\cos E} + \frac{e\sin E dE(\cos E - e)}{(1-e\cos E)^2}. \qquad (4.65)$$

We isolate $d\nu$ and reorganize using a common denominator to get

$$\begin{aligned}
d\nu &= \frac{1-e\cos E+e\cos E-e^2}{(1-e\cos E)^2}\frac{\sin E}{\sin\nu}dE \\
&= \frac{1-e^2}{1-e\cos E}\frac{1}{1-e\cos E}\frac{\sin E}{\sin\nu}dE \\
&= \frac{p/r}{r/a}\frac{\sin E}{\sin\nu}dE = \frac{a\sqrt{1-e^2}}{r}dE. \qquad (4.66)
\end{aligned}$$

We now have what we need to make the change of variable,

$$\begin{aligned}
h(t-T) = \int_0^\nu r^2 d\nu &= \int_0^E r^2\frac{a\sqrt{1-e^2}}{r}dE \\
&= \frac{p}{\sqrt{1-e^2}}\int_0^E r dE \\
&= \frac{pa}{\sqrt{1-e^2}}\int_0^E (1-e\cos E)dE \\
&= \frac{pa}{\sqrt{1-e^2}}(E-e\sin E). \qquad (4.67)
\end{aligned}$$

Using $h=\sqrt{\mu p}$ gives the final result

$$t-T = \sqrt{\frac{a^3}{\mu}}\big(E-e\sin E\big), \qquad (4.68)$$

which is identical to the geometric result.

4.3 Parabolic Time of Flight as a Function of D

Parabolic time of flight is most easily developed using the analytical method. Gravity is a central force no matter what the shape of the orbit, so angular momentum conservation applies to all the conic-section orbits. We will use this to develop the equivalent of a time-of-flight formula for parabolic motion. We start by recognizing that

$$h = r^2\dot{\nu} = \sqrt{\mu p}, \tag{4.69}$$

and that for a parabola $e = 1$, so

$$r = \frac{p}{1+\cos\nu}. \tag{4.70}$$

Together these give

$$\sqrt{\mu p} = \left(\frac{p}{1+\cos\nu}\right)^2 \dot{\nu}, \tag{4.71}$$

which gives

$$\frac{\sqrt{\mu p}}{p^2} = \frac{1}{p}\sqrt{\frac{\mu}{p}} = \frac{\dot{\nu}}{\left(1+\cos\nu\right)^2}. \tag{4.72}$$

We remind ourselves that

$$\dot{\nu} = \frac{d\nu}{dt}, \tag{4.73}$$

so that

$$\frac{1}{p}\sqrt{\frac{\mu}{p}}dt = \frac{d\nu}{\left(1+\cos\nu\right)^2}. \tag{4.74}$$

We integrate both sides to get

$$\frac{1}{p}\sqrt{\frac{\mu}{p}}\int_T^t dt = \int_0^\nu \frac{d\nu}{\left(1+\cos\nu\right)^2}. \tag{4.75}$$

This allows us to calculate the time of flight from periapsis, $\nu = 0$ at time T, to some angular position ν at time t, or, equivalently, the time to periapsis from some angle ν. The indefinite form of this integral is found in integral tables, in this case I found it on line at

www.sosmath.com/tables/integral/integ20/integ20.html with the result

$$\int \frac{d\nu}{\left(1+\cos\nu\right)^2} = \frac{1}{2}\tan\frac{\nu}{2} + \frac{1}{6}\tan^3\frac{\nu}{2}. \tag{4.76}$$

Thus,

$$\frac{1}{p}\sqrt{\frac{\mu}{p}}\left(t-T\right) = \frac{1}{2}\left(\tan\frac{\nu}{2} + \frac{1}{3}\tan^3\frac{\nu}{2}\right). \tag{4.77}$$

This can be rearranged to give

$$\begin{aligned} t-T &= p\sqrt{\frac{p}{\mu}}\frac{1}{2}\left(\tan\frac{\nu}{2} + \frac{1}{3}\tan^3\frac{\nu}{2}\right) \\ &= \frac{1}{2\sqrt{\mu}}\left(p^{3/2}\tan\frac{\nu}{2} + \frac{p^{3/2}}{3}\tan^3\frac{\nu}{2}\right) \\ &= \frac{1}{2\sqrt{\mu}}\left(pD + \frac{1}{3}D^3\right), \end{aligned} \tag{4.78}$$

where $D = \sqrt{p}\tan\frac{\nu}{2}$. In this form it is called Barker's equation, and looks very similar to Kepler's equation. For this reason D is called the parabolic eccentric anomaly. This equation is transcendental.

Following the practice of the text, this equation allows us to describe the motion between two arbitrary points in the parabolic orbit as

$$t - t_o = \frac{1}{2\sqrt{\mu}}\left[\left(pD + \frac{1}{3}D^3\right) - \left(pD_o + \frac{1}{3}D_o^3\right)\right]. \tag{4.79}$$

Some authors choose to identify the leading factor of

$$\sqrt{\frac{p^3}{\mu}} \tag{4.80}$$

in the next-to-last step as the equivalent of the term

$$\sqrt{\frac{a^3}{\mu}} \tag{4.81}$$

in Kepler's equation, so that it can be used to define a parabolic period as

$$T_p = 2\pi\sqrt{\frac{p^3}{\mu}}. \tag{4.82}$$

This is a definition of convenience only. The parabolic orbit does not repeat, so it has no true period.

4.4 Hyperbolic Time of Flight as a Function of F

The hyperbola is probably the least familiar of the conic sections, and hyperbolic orbits are probably the least studied orbits. Hyperbolic orbits are of great importance to the exploration of outer space - if we are ever to travel to another star we will have to take a hyperbolic orbit out of the Solar System.

The definition of the hyperbolic time of flight proceeds similarly to that of parabolic time of flight, starting with conservation of angular momentum,

$$h = r^2 \frac{d\nu}{dt}. \tag{4.83}$$

Substituting, separating, and integrating gives

$$h(t - T) = \int_0^\nu \frac{d\nu}{(1 + e\cos\nu)^2}. \tag{4.84}$$

This integral is not as easy as the previous ones, and success will involve defining the hyperbolic sine and cosine in analogy with the familiar trigonometric functions on circles. This will require a new definition of angle in terms of area.

4.4.1 Area as a Measure of Angle

Kepler's equation is solved using angles in radian measure. This is because radians are the correct, dimensionless measure of angle for terms like $E - e\sin E$, where both E and e are dimensionless. Radians are measured using the ratio of arc length to radius, giving the desired dimensionless quantity, and are especially easy to measure on the unit circle. We have seen that it is difficult to measure arc length on an ellipse, and we accept that it is similarly difficult on a hyperbola.

Let's develop area as a measure of angle starting with the unit circle and obtaining the familiar result for trig functions. The following material draws on the excellent web page on hyperbolas by Dr. James B. Calvert at the University of Denver,
http://mysite.du.edu/ jcalvert/math/hyperb.htm
but note that his η and xi axes are interchanged in the figure labeled "Finding the Area A." The general properties of the equilateral or right hyperbola are given at

http://mathworld.wolfram.com/RectangularHyperbola.html
which we cite as
Weisstein, Eric W. "Rectangular Hyperbola." From *Mathworld* – a Wolfram Web Resource.
http://mathworld.wolfram.com/RectangularHyperbola.html

There is valuable material on hyperbolas from Wikipedia at
http://en.wikipedia.org/wiki/Hyperbolic_function

which we cite as

Wikipedia contributors, "Hyperbolic function," Wikipedia, The Free Encyclopedia,

http://en.wikipedia.org/w/index.php?title=Hyperbolic_function&oldid=156619267 (accessed September 14, 2007).

There is an excellent animation of the area measure of angle on a hyperbola at
http://en.wikipedia.org/wiki/Image:HyperbolicAnimation.gif
which does not contain citation instructions in its toolbox, so we cite it as

Wikipedia contributors, "Image:HyperbolicAnimation.gif," Wikipedia, The Free Encyclopedia,

http://en.wikipedia.org/wiki/Image:HyperbolicAnimation.gif (accessed September 30, 2007).

We consider a radius vector on the unit circle that is allowed to sweep out area as it moves from its reference position, $\theta = 0$ on the x axis, to some arbitrary position θ. in moving through a small angle $d\theta$ it sweeps out a skinny triangle whose area is $\frac{1}{2}bh$, where b is the base nd h is the height. The base is just the radius, r, and the height is $rd\theta$, so

$$dA = \frac{1}{2}bh = \frac{1}{2}r^2 d\theta. \tag{4.85}$$

Integrating from $\theta = 0$ to θ gives

$$A = \frac{1}{2}r^2\theta. \tag{4.86}$$

On the unit circle $r = 1$ and $A = \frac{\theta}{2}$, or $2A = \theta$. On any other circle the same result can be obtained by dividing by r^2. If we set $t = 2A$ then the familiar trig functions, $\cos t$, $\sin t, \tan t$, and $\cot t$ are shown in the figure.

We now pursue the same thing using a hyperbola. First, we need the hyperbolic equivalent of a unit circle. Given the general formula for an ellipse as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{4.87}$$

and the unit circle as a special case of an ellipse with $a = b = 1$, we get

$$x^2 + y^2 = 1. \tag{4.88}$$

Now we write the general formula for a hyperbola as

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{4.89}$$

and again choose $a = b = 1$ as a special case to get

$$x^2 - y^2 = 1. \tag{4.90}$$

We are tempted to call this the unit hyperbola, but it is already known as the equilateral or right hyperbola. It is special because its asymptotes are perpendicular, hence the name right hyperbola. It is known to have $e = \sqrt{2}$. It just touches, or osculates, the unit circle, sharing the single point $(x, y) = (1, 0)$.

We will find that areas associated with the unit hyperbola are more easily studied if we rotate the coordinates and use the asymptotes as the axes. Following Calvert, we define

$$x - y = \eta\sqrt{2} \tag{4.91}$$

and

$$x + y = \xi\sqrt{2}, \tag{4.92}$$

so the parabola becomes

$$1 = x^2 - y^2 = (x - y)(x + y) = 2\eta\xi, \tag{4.93}$$

or

$$\eta\xi = \frac{1}{2}. \tag{4.94}$$

An equilateral hyperbola is graphed in the accompanying figure, adapted from Calvert, which also shows the x, y, η, and ξ axes. We consider a radius vector from $\mathcal{O}$, the common origin of the $x - y$ and $\eta - \xi$ coordinates, to the hyperbola, starting at pericenter, which is the symmetry point labeled

A. Allow an orbiting object to proceed along the hyperbola to a point P and allow the radius vector to sweep out area during the motion. The hyperbolic sine, cosine, and tangent are shown, along with the swept area. $\widehat{\mathcal{O}AP}$ is called a hyperbolic triangle, and we wish to know its area in order to define the area measure of angle. The shape of the hyperbolic triangle is unfamiliar and complex to integrate. Instead of a brute-force integration we seek a simpler method.

We begin by drawing the perpendicular from A to the ξ axis, and calling the intersection B. Similarly, we draw the perpendicular from P to the ξ axis and call the intersection Q. The $x-y$ coordinates of point A are $(1,0)$, so its $\eta-\xi$ coordinates are $(1/\sqrt{2}, 1/\sqrt{2})$, and those of point B are $(0, 1/\sqrt{2})$. Point P is at an arbitrary point (x,y) or (η,ξ) on the hyperbola, so point Q has coordinates $(0,\xi)$, and the distance between P and Q is ξ.

We seek the area of hyperbolic triangle $\widehat{\mathcal{O}AP}$, which is equal to the area of figure $\mathcal{O}APQ$ minus the area of triangle $\widehat{\mathcal{O}PQ}$. The area of the triangle is relatively easy, so let's see if we can calculate it. Let the triangle's base be the segment $\mathcal{O}P$ and the height be the segment PQ. Then

$$\frac{1}{2}bh = \frac{1}{2}\eta\xi = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}. \tag{4.95}$$

The closed-form result, $\frac{1}{2}\eta\xi$, looks like a general result, so let's calculate the area of triangle $\widehat{\mathcal{O}AB}$. Let its base be segment $\mathcal{O}B$ and its height be segment AB. Then

$$\frac{1}{2}bh = \frac{1}{2}\cdot\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}} = \frac{1}{4}. \tag{4.96}$$

The areas are the same! This result turns out to be general, but, more importantly, it means that the area of the hyperbolic triangle $\widehat{\mathcal{O}AP}$ is equal to the area of figure $\mathcal{O}APQ$ minus the area of triangle $\widehat{\mathcal{O}AB}$, which is the area of figure $ABQP$. This is a figure whose area should be easy to obtain by integration, so let's try. Divide the figure into strips perpendicular to the ξ axis of width $d\xi$. Each will have height perpendicular to the ξ axis of η, so the strip will have a small area

$$dA = \eta d\xi = \frac{d\xi}{2\xi}. \tag{4.97}$$

We already know the positions of points B and Q in $\eta-\xi$ coordinates as $(0, 1/\sqrt{2})$ and $(0,\xi)$, respectively. These easily allow us to specify the limits of integration to find the area, if we note that $\xi = \frac{x+y}{\sqrt{2}}$. Then

$$A = \int_{\frac{1}{\sqrt{2}}}^{\frac{x+y}{\sqrt{2}}} \frac{d\xi}{2\xi} = \frac{1}{2} \ln{(x+y)}. \tag{4.98}$$

In analogy with the area definition of angle for the trig functions, the area that we need for the hyperbolic functions is

$$t = 2A = \ln{(x+y)}. \tag{4.99}$$

The next figure shows the trig functions defined on the unit circle and the hyperbolic functions defined on the equilateral hyperbola. We note that t is not literally an angle, but, rather, an area equivalent of angle appropriate to the equilateral hyperbola.

4.4.2 Hyperbolic Time of Flight

Figure 4.2-5 of our text contains the cryptic equation

$$F = \frac{area\ QOV}{\frac{1}{2}a^2}. \tag{4.100}$$

Now we know why. This is the area measure of angle applied to the equilateral hyperbola.

Now we need to generalize our approach, for we expect that very few objects will be in orbits that are equilateral hyperbolas. Figure 4.2-5 shows a general hyperbola drawn with an equilateral hyperbola as an auxiliary figure, to be used in the same way that the circle was for the general ellipse. The generalized equilateral hyperbola is one whose asymptotes are perpendicular, but does not necessarily have $a = 1$. It is described by

$$\frac{x_e^2}{a^2} - \frac{y_e^2}{a^2} = 1, \tag{4.101}$$

or

$$x_e^2 - y_e^2 = a^2, \tag{4.102}$$

where the subscript e indicates equilateral. We note that switching to this generalized equilateral hyperbola will increase the area of hyperbolic triangles by a factor of a^2.

A general parabola that osculates the equilateral one at $(x, y) = (\pm a, 0)$ is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{4.103}$$

In analogy with the analysis of elliptical motion, we seek a relation between the y coordinates of these two hyperbolas for a fixed value of x. Remembering that $b^2 = a^2(e^2 - 1)$, we see

$$\frac{y_e^2}{y^2} = \frac{x^2 - a^2}{\frac{b^2 x^2}{a^2} - b^2} = \frac{x^2 - a^2}{a^2(e^2-1)(\frac{x^2}{a^2} - 1)} = \frac{1}{e^2 - 1}, \tag{4.104}$$

or

$$\frac{y_e}{y} = \pm\frac{1}{\sqrt{e^2 - 1}}. \tag{4.105}$$

It is helpful to remember at this point that the convention used in our text is that $e > 1$.

We can now define the hyperbolic anomaly, F, using Figure 4.2-5 of the text, and derive the time of flight for the hyperbolic orbit. The derivation is missing from our text, so the derivation below comes from "Fundamentals of Astrodynamics and Applications," third edition, by David A. Vallado 2007, published jointly by Microcosm Press, Hawthorne, CA, and Springer, New York. We note that the standard notation for the hyperbolic anomaly is F, but Vallado uses H. We see

$$\sinh F = \frac{-r \sin \nu}{a\sqrt{e^2 - 1}}, \tag{4.106}$$

and

$$\cosh F = \frac{ae + r \cos \nu}{a}. \tag{4.107}$$

As in the elliptical case, we seek a relation between F and ν, so replace r with its value from the trajectory equation,

$$r = \frac{p}{1 + e\cos\nu} = \frac{a(1 - e^2)}{1 + e\cos\nu}, \tag{4.108}$$

to get

$$\sinh F = \frac{-a(1 - e^2)\sin\nu}{a(1 + e\cos\nu)\sqrt{e^2 - 1}} = \frac{\sqrt{e^2 - 1}\sin\nu}{1 + e\cos\nu}, \tag{4.109}$$

and

$$\cosh F = e + \frac{1-e^2}{1+e\cos\nu}\cos\nu = \frac{e(1+e\cos\nu)+(1-e^2)\cos\nu}{1+e\cos\nu} = \frac{e+\cos\nu}{1+e\cos\nu}. \tag{4.110}$$

Here we have derived both hyperbolic functions, but have a preference for the hyperbolic sine because it is not double valued.

We will also want a way to calculate ν from F, which we can get by inverting the equation for hyperbolic cosine that we just derived,

$$(1+e\cos\nu)\cosh F = e+\cos\nu, \tag{4.111}$$

so

$$\cosh F + (e\cosh f - 1)\cos\nu = e, \tag{4.112}$$

and

$$\cos\nu = \frac{e-\cosh F}{e\cosh F - 1} = \frac{\cosh F - e}{1-e\cosh F}. \tag{4.113}$$

Similarly, we want an expression for r in terms of F, which we get by solving for r in our original equation for $\cosh F$,

$$\cosh F = \frac{ae + r\cos\nu}{a}, \tag{4.114}$$

so

$$r = \frac{a\cosh F - ae}{\cos\nu}. \tag{4.115}$$

substituting for $\cos\nu$ with its equivalent in terms of F gives

$$r = \frac{(a\cosh F - ae)(1-e\cosh F)}{\cosh F - e}, \tag{4.116}$$

and simplifying gives the desired result

$$r = a(1-e\cosh F). \tag{4.117}$$

We can also solve our original equation for $\sinh F$ to calculate $\sin\nu$ from F,

$$\sin\nu = \frac{-a\sinh F\sqrt{e^2-1}}{r} = \frac{-a\sinh F\sqrt{e^2-1}}{a(1-e\cosh F} = \frac{-\sinh F\sqrt{e^2-1}}{1-e\cosh F}. \tag{4.118}$$

In the derivation for elliptical motion we next differentiated the expression for $\cos\nu$ to get a relation between $d\nu$ and dE. We follow the same procedure, and differentiate the expression for $\cos nu$ to get a relation between $d\nu$ and dF,

$$\begin{aligned} -\sin\nu d\nu &= \frac{\sinh F}{1-e\cosh F}dF - \frac{\cosh F - e}{(1-e\cosh F)^2}(-e\sinh F)dF \\ &= \frac{\sinh F(1-e\cosh F)+(\cosh F - e)e\sinh F}{(1-e\cosh F)^2}dF, \end{aligned} \tag{4.119}$$

so

$$d\nu = \frac{\sinh F(1-e^2)}{-\sin\nu(1-e\cosh F)^2}dF. \tag{4.120}$$

We replace $\sin\nu$ with its value in terms of F and note that we already know

$$r = a(1-e\cosh F), \tag{4.121}$$

so

$$(1-e\cosh F)^2 = \frac{r^2}{a^2}, \tag{4.122}$$

so

$$d\nu = \frac{\sinh F(1-e^2)}{\frac{a\sinh F\sqrt{e^2-1}}{r}\frac{r^2}{a^2}}dF. \tag{4.123}$$

This simplifies to

$$d\nu = \frac{a(1-e^2)}{r\sqrt{e^2-1}}dF = \frac{-a\sqrt{e^2-1}}{r}dF. \tag{4.124}$$

We can now go back to angular momentum conservation with a full hyperbolic toolbox. Reminding ourselves that

$$h = r^2\frac{d\nu}{dt}, \tag{4.125}$$

so that

$$hdt = r^2 d\nu, \tag{4.126}$$

and

$$h\int_T^t dt = \int_0^\nu r^2 d\nu, \tag{4.127}$$

so

$$h(t-T) = \int_0^\nu r^2 d\nu. \tag{4.128}$$

Our hyperbolic toolbox allows us to change the variable of integration from ν to F,

$$h(t-T) = \int_0^F r^2 \frac{-a\sqrt{e^2-1}}{r} dF = -a\sqrt{e^2-1}\int_0^F r dF. \tag{4.129}$$

We know that $r = a(1 - e\cosh F)$ and $h = \sqrt{\mu p}$, so

$$t - T = \frac{-a^2\sqrt{e^2-1}}{\sqrt{\mu p}} \int_0^F (1 - e\cosh F) dF. \tag{4.130}$$

Happily, we can integrate by inspection, replace p, and reorganize to get

$$t - T = \sqrt{\frac{-a^3}{\mu}}(e\sinh F - F). \tag{4.131}$$

This allows the time of flight between two arbitrary points in the hyperbola to be calculated.

Chapter 5

Orbit Determination from Two Positions and Time

Following the authors's suggestions for a one-semester course, we will skip this chapter.

Chapter 6

Ballistic Missile Trajectories

6.1 History

Here are a few points to consider:

1. The Treaty of Versailles, that ended World War I, forbade the Germans from developing long-range artillery. They heeded this, and developed long-range rockets, or missiles, instead.

2. Robert Goddard, an American, was also developing rockets. He chose to do this in New Mexico, in the desert Southwest. His development of liquid fuel probably was known to the Germans and copied by them.

3. The German V-1 and V-2 rockets were successful as rockets, but not as weapons. They came close to succeeding, and might have if they had been more plentiful and deployed earlier.

4. The United States captured many V-2 rockets, parts, and personnel when Germany was defeated at the end of World War II. The people and material goods were transported to New Mexico later in 1945. The purpose was to develop a rocket program.

5. A large section of New Mexico had already been claimed by the government as the Trinity Site, to test the atomic bomb. This was set up for missile testing, and became White Sands Proving Ground, later White Sands Missile Range.

6.2 Purpose

This is our first exercise in mission planning. We will drop the usual homework and make mission planning reports our focus and purpose. Assume that it is November, 1945, and you are part of the team developing a missile program. How large a test range is needed?

For a summary of the early efforts at White Sands, please see the web sites
http://www.wsmr.army.mil/pao/FactSheets/V2/v-2.htm
http://www.wsmr.army.mil/pao/FactSheets/V2/v-2tab.htm
We are to make a first report that develops the theory of suborbital rocket flights, makes a recommendation about how large the missile range has to be, and indicates what accuracy is needed in guiding a missile so that it makes a successful flight and lands back on the missile range. The missiles must be used to gain knowledge of missile flight and the effects of the atmosphere on suborbital missiles.

6.3 Basics of Ballistic Missiles

The trajectory of a ballistic missile has three parts: powered flight, free flight, and re-entry. During powered flight and re-entry there are continuous external forces other than gravity on the missile, so these parts of the flight cannot be analyzed with two-body mechanics. There are entire courses on these segments of the flight, so we will ignore them, and concentrate on free flight, which can be analyzed with the approach that we have developed.

We will begin by ignoring the rotation of the Earth, with the goal of including rotation later. We will also assume that the free-flight trajectory is symmetrical.

We want to relate the range of a missile to its orbital properties at burnout, the instant at which the rocket engine stops firing. Our first practical question is, "Given the position and velocity at burnout, what flight-path angle at burnout of powered flight is required?" We will use the symbols r_{bo} and v_{bo}. We will describe the trajectory with the powered-flight range angle, Γ, the free-flight range angle, Ψ, the re-entry range angle, Ω, and the total range angle, Λ, such that

$$\Lambda = \Gamma + \Psi + \Omega. \tag{6.1}$$

6.3.1 The Non-dimensional Parameter, Q

It is convenient to define

$$Q \equiv \left(\frac{v}{v_{cs}}\right)^2 = \frac{v^2 r}{\mu}. \tag{6.2}$$

Q is the square of the ratio of the current speed to the local circular speed. If $Q = 1$ the object has the local circular speed. This does not guarantee that it is moving in a circle, for the direction may be incorrect. We have already seen that this is the case on the minor axis of an ellipse. If $Q = 2$ the object has escape speed, and if $Q > 2$ the object has hyperbolic speed.

We can use the energy equation,

$$\mathcal{E} = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}, \tag{6.3}$$

and substitute $\frac{\mu Q}{r}$ for v^2 to get

$$Q = 2 - \frac{r}{a}, \tag{6.4}$$

or

$$a = \frac{r}{2 - Q}. \tag{6.5}$$

6.3.2 The Free-Flight Range Equation

We can apply the trajectory equation to burnout to get

$$r_{bo} = \frac{p}{1 + e\cos\nu_{bo}}. \tag{6.6}$$

Solving for $\cos\nu_{bo}$ gives

$$\cos\nu_{bo} = \frac{p - r_{bo}}{e r_{bo}}. \tag{6.7}$$

The free-flight trajectory is assumed to be symmetrical, so half of the free-flight range angle lies on each side of the major axis, and

$$\cos\frac{\Psi}{2} = \cos\left(180^o - \nu_{bo}\right) = -\cos\nu_{bo}. \tag{6.8}$$

This allows us to write

$$\cos\frac{\Psi}{2} = \frac{r_{bo} - p}{e r_{bo}} = \frac{1 - p/r_{bo}}{e}. \tag{6.9}$$

We can make this more useful by writing

$$p = \frac{h^2}{\mu} = \frac{r^2 v^2 \cos^2 \phi}{\mu} = rQ \cos^2 \phi. \tag{6.10}$$

We remember $p = a(1 - e^2)$, so

$$e^2 = 1 - \frac{p}{a}. \tag{6.11}$$

Substituting $p = rQ \cos^2 \phi$ and $a = \frac{r}{2-Q}$ gives

$$e^2 = 1 + Q(Q - 2) \cos^2 \phi. \tag{6.12}$$

The eccentricity and parameter are constants, so their values can be calculated at burnout, or anywhere else on a given orbit. Thus,

$$e^2 = 1 + Q_{bo}(Q_{bo} - 2) \cos^2 \phi_{bo}, \tag{6.13}$$

and

$$p = r_{bo} Q_{bo} \cos^2 \phi_{bo} \tag{6.14}$$

may be substituted into the expression for the cosine of $\frac{\Psi}{2}$ to get

$$\cos \frac{\Psi}{2} = \frac{1 - Q_{bo} \cos^2 \phi_{bo}}{\sqrt{1 + Q_{bo}(Q_{bo} - 2) \cos^2 \phi_{bo}}}. \tag{6.15}$$

We can now calculate the free-flight range angle from the burnout values of r, v, and ϕ.

6.3.3 The Flight-Path Angle equation

If we could calculate ϕ_{bo} needed to have a missile hit a target given r_{bo} and v_{bo} that would be even more useful. That is our next goal.

To do this we make use of the optical properties of an elliptical mirror. Light emitted at one focus of the ellipse is reflected by the surface of the ellipse to the other focus. The angle of incidence equals the angle of reflection, so this means that the normal to the ellipse bisects the angle between the light paths, and the angle of incidence and the angle of reflection equal ϕ_{bo}. We call the radius from the occupied focus of an elliptical orbit to the burnout point r_{bo}, and the associated radius from the unoccupied focus to the burnout point r'_{bo}. Then

$$r_{bo}\sin\frac{\Psi}{2} = r'_{bo}\sin\left(180^o - 2\phi_{bo} + \frac{\Psi}{2}\right). \tag{6.16}$$

Reorganizing gives the result,

$$\sin\left(2\phi_{bo} + \frac{\Psi}{2}\right) = \frac{r_{bo}}{r'_{bo}}\sin\frac{\Psi}{2}, \tag{6.17}$$

and substituting $r_{bo} = a(2 - Q_{bo})$ and $r_{bo} + r'_{bo} = 2a$ gives the flight-path angle equation,

$$\sin\left(2\phi_{bo} + \frac{\Psi}{2}\right) = \frac{2 - Q_{bo}}{Q_{bo}}\sin\frac{\Psi}{2}. \tag{6.18}$$

In using this equation we take Ψ as known and ϕ_{bo} as unknown. The equation then gives a number equal to $\sin(2\phi_{bo} + \frac{\Psi}{2})$. In general there will be two angles with the calculated sine, so the equation has two solutions. Following the example in the text, take $\Psi = 90^o$ and $Q_{bo} = 0.9$. Then

$$\sin(2\phi_{bo} + 45^o) = \frac{2 - 0.9}{0.9}\sin 45^o = 0.864 \simeq 0.866. \tag{6.19}$$

The sine is close enough to $\frac{\sqrt{3}}{2}$ that the authors of our text choose equality. There are two angles with the appropriate sine, 60^o and 120^o, so there are two values of the flight-path angle,

$$\phi_{bo} = 7.5^o, 37.5^o. \tag{6.20}$$

These are called the low and high trajectories, respectively. This is closely related to the fact that when throwing a ball at a fixed speed on a level playing field there are two trajectories to hit a target.

The nature of the trajectories depends on the value of Q_{bo}. If $Q_{bo} < 1$ then there will be a maximum value of Ψ that causes the right-hand side of the flight-path equation to equal one. This establishes the maximum range for a missile with Q_{bo} less than one. The maximum range angle will always be less than 180^o for Q_{bo} less than one, so if Ψ is attainable there will be a low and a high trajectory.

If $Q_{bo} = 1$ then there will be trajectory that is a circular orbit joining the launch and target points. This circular orbit is likely to be highly impractical because most or all of it is in the Earth's atmosphere.

If $Q_{bo} > 1$ then there will be one positive and one negative value for ϕ_{bo}. The negative value is impractical, even for the School of Mines, because it represents an orbit within the Earth. Only the high trajectory is practical. When $Q_{bo} > 1$ is is possible to have a range angle in excess of 180^o.

Chapter 7

Lunar Trajectories

7.1 Sphere of Influence

We have decided to skip the study of lunar trajectories in favor of interplanetary ones. We do need to cover the topic of spheres of influence, which is an outgrowth of the restricted three-body problem, and the closely associated topic of patched-conic approximations.

After the success of the two-body problem much effort was devoted to studying the three body problem. There is no general, closed-form solution for the motion of three bodies moving only under their mutual gravity, and it looks as though none exists. Many helpful results can be found on the restricted three-body problem, in which one of the objects is massive, one is much less massive, and one has a mass so small that can be neglected. An example would be the Sun, a planet, and an artificial satellite. Under these conditions the Sun's gravity dominates the entire Solar System, but there is a limited volume around the planet in which a satellite orbit may be thought of as dominated by the planet's gravity with the Sun's gravity providing a perturbation. This was worked on by Laplace and his successors throughout the nineteenth century. I have tried to obtain copies of the original work, but so far do not have any. Therefore, we will have to accept a statement of the results and hope that the details can be filled in later.

The result is that the sphere of influence of the planet is approximated by a sphere of radius

$$r_s = r\left(\frac{m_p}{m_s}\right)^{\frac{2}{5}}, \tag{7.1}$$

where r_s is the radius of the sphere of influence, r is the distance between the planet and the Sun, m_p is the mass of the planet, and m_s is the mass of the Sun.

7.2 The Patched Conic Approximation

We now assume that a spacecraft that moves from the inside to the outside of the planet's sphere of influence, or from outside to inside, simply follows two conic section orbits, one inside and one outside. This is overly simple, but allows a closed-form solution for the overall motion that can serve as an initial model for a fully numerical solution to the overall motion.

Chapter 8

Interplanetary Trajectories