

A GLOBAL POSITIONING SYSTEM (GPS) PRIMER

by

Robert Augustson

B.A. Regents College, 2000

A thesis submitted to the
University of Colorado at Denver
in partial fulfillment
of the requirements for the degree of
Master of Science
Applied Mathematics
2003

This thesis for the Master of Science

degree by

Robert Augustson

has been approved

by

William Briggs

Lynn Bennethum

Lynn Johnson

Date

Augustson, Robert A. (M.S., Applied Mathematics)
A Global Positioning System (GPS) Primer
Thesis directed by Professor William Briggs

ABSTRACT

The Global Positioning System (GPS) has been designed to enable one to determine one's position in a latitude - longitude reference frame on or near the Earth to within several meters. Given that the Earth is an oblate spheroid (that is, ellipsoidal), a reference ellipsoid must be defined along with a latitude - longitude convention. The GPS uses a constellation of artificial satellites orbiting the Earth and the measurement of the distances from an observer to four satellites to determine position and time, x, y, z, t , by trilateration. Distance measurement is accomplished by measuring the propagation time delay of radio waves transmitted by a satellite and received by an observer. Given that electromagnetic waves propagate at approximately 300,000,000 meters per second, a differential time delay of 10 nanoseconds represents a differential distance of approximately 3 meters, thus time delays must be measured within a few nanoseconds. To accurately determine the position of an observer with respect to several orbiting satellites one must determine the orbits of the satellites, then determine the location of each satellite in its orbit within a few meters with respect to the Earth's center of mass at a particular instant in time. The absolute time used to determine orbital position must therefore be known within nanoseconds, and the observer's clock must be synchronized with those of the satellites used for the determination of position. Absolute time is therefore a fourth unknown, which is determined by distance to the fourth satellite. Since orbital positions are determined with respect to the Earth's center of mass, and therefore the observer's position can be determined with respect to the Earth's center of mass, a final conversion must be made relating the observer's position with respect to the Earth's center of mass and the latitude - longitude convention of the reference ellipsoid. This thesis is an exposition of the mathematical calculations and methods involved in the GPS.

This abstract accurately represents the content of the candidate's thesis. I recommend its publication.

signed _____

CONTENTS

Figures	vi
---------------	----

Chapter

1.	Introduction	1
1.1	The Reference Ellipsoid	3
2.	Using Artificial Satellites For Navigation	10
2.1	Satellite Orbits	11
2.1.1	Newton's Laws	11
2.1.2	The Two-Body Problem	12
2.1.3	The Ellipse	19
2.2	The Elliptical Orbit	22
2.2.1	Earth-Centered Earth-Fixed Coordinate System	26
2.2.2	Mean and Eccentric Anomaly	28
2.2.3	Perturbed Orbits	32
3.	The Global Positioning System	40
3.1	The Space Segment/User Segment Interface	41
3.1.1	GPS Satellite Signals	41
3.1.2	The NAV Message	44
3.2	The User Segment	45
3.2.1	Calculating Delays	46
3.2.2	Ionospheric Delay Model	48
3.2.3	Calculating SV Position	50
3.2.4	Calculating User Position	52
3.2.5	Dilution of Precision	57
4.	Final Thoughts	58

Appendix

A.	Ephemeris Definitions	59
B.	Subframes 1, 2 and 3	60
C.	ECEF Algorithm And HOW	63

<u>References</u>	65
-------------------------	----

FIGURES

1.1	Latitude.	4
1.2	Ellipsoidal Coordinates.	6
1.3	Conversion From Cartesian To Ellipsoidal Coordinates.	8
2.1	Earth And Satellite Position Vectors.	13
2.2	Satellite Polar Vector, Earth Origin.	15
2.3	The Elliptical Locus.	20
2.4	The Orbital Plane.	23
2.5	Earth Centered, Earth-Fixed Coordinates.	26
2.6	Eccentric Anomaly.	29
2.7	Potential Referenced To A Planet's Center Of Mass.	33
3.1	Significant Subframe Contents.	44
3.2	Application Of Correction Parameters.	47
3.3	Receiver Clock Offset.	55

1 Introduction

The Greek scholars at Alexandria realized that the Earth is spherical in shape. Indeed, in the fourth century B.C.E. Aristotle observed in his book *On the Heavens* that the shadow of the Earth on the moon during a lunar eclipse is curved, and also that the elevation angle of certain stars changes as one travels north and south. Both arguments correctly identify a spherically shaped Earth.

Euclid's geometry was applied to measurements of land and of the Earth and Eratosthenes, the librarian at Alexandria about 250 B.C.E., made a calculation of the circumference of the Earth within one percent [9]. Eratosthenes used the sun angles at local noon on the summer solstice at two locations on a north-south line, Syene (now Aswan, Egypt) where the sun was directly overhead at noon on the summer solstice, and Alexandria, 500 miles due north of Syene; he then used simple geometry to deduce his result.

Claudius Ptolemy (c. 100-178 C.E.) developed plane and spherical trigonometric tables and a complete mathematical description of astronomy as then known by the Greeks in his *Mathematical Collection*, a work composed of thirteen books. Islamic scholars later referred to Ptolemy's books as *al-magisti*, "the greatest," and this work became known as the *Almagest*, a work unsurpassed until the sixteenth century. Ptolemy, observing the shape of the Mediterranean Sea, introduced the terms latitude (across) and longitude (in the long direction) [16].

He made observations of the heavens near Alexandria, and in another book, *Geography*, he compiled the latitudes and longitudes of places in the known world. In the *Geography* he also discussed the projections needed for map making [10].

The *Almagest* contains three plane projective maps of the world as it was then known. These early maps attest to a desire to know one's position relative to known landmarks, and by extension, to know how to navigate to other locations whose positions are known with respect to known landmarks. These early maps are products of triangulation and astronomical observations, and are limited by the accuracy of the trigonometric tables and the devices used to measure distances and angles.

More accurate measurements and calculations of distances and angles improved triangulation methods and improved the accuracy of maps and land navigation. Astronomical observations could provide one's latitude, for example, by measuring sun angles at local noon and comparing with tabulated data showing sun angles at noon at different latitudes versus the day of the year. One

could also make astronomical observations at night, measuring the elevation angles of known stars and comparing with the elevations of the same stars at locations of known latitude. While nowadays measuring the elevation angle of Polaris, the north star, yields one's approximate latitude, this has not always been the case. Due to precession, true north, the direction of the Earth's axis, describes a circle in the heavens with a period of about 25,800 years (a Plutonic Year). It was not until Christopher Columbus' time that Polaris could be used as a north star. About 5000 years ago, when the pyramids were built, the star Thuban, in the constellation Draco, was the "north star," and the star Vega will be near the celestial north pole some 13,000 years from now.

Longitude was determined by the Greek mathematicians by application of geometry along with knowledge of the distances and directions between landmarks of known latitude and longitude. Determining distances and directions to known landmarks over large bodies of water and out of sight of land prevented knowing one's longitude exactly while at sea on a ship; therefore navigation over water was difficult and the accurate determination of one's position was impossible. To determine longitude, one needs to know at local noon (or some other standard time) what angle the Earth has rotated through since noon (or other standard time) at a landmark of known longitude; therefore one must have a time standard to know how much time has elapsed between noon at the landmark of known longitude and local noon for such determination.

Travel at sea was therefore a perilous experience before modern navigation methods. Navigation was by dead reckoning - by estimating speed and using a compass or the sun and stars to provide direction and knowledge of latitude. Longitude information was provided occasionally by observations of known landmarks. Both ships and passengers were imperiled when weather obscured observations. After experiencing naval disasters due to the lack of accurate determination of position at sea, the British government in 1714 offered a prize of £20,000 for the development of a method for determining longitude at sea within half a degree.

Galileo Galilei's book *The Starry Messenger* published in Venice in 1610 revealed telescopic discoveries including the four inner satellites of Jupiter. Galileo suggested using these satellites as a celestial clock, and James Bradley, the third Astronomer Royal of England successfully applied a knowledge of the motion of these satellites in 1726 to determine the longitude of Lisbon and New York with considerable accuracy. The difficulty of observing these satellites from the deck of a ship, and the difficulty of producing an accurate ephemeris of Jupiter's satellites made this method impractical for use at sea.

Observation of the predictable motions of the moon, the Earth's natural satellite, can provide timing measurements. With an accurate ephemeris of the moon, observations of the moon can yield one's longitude. Tobias Mayer, a

German cartographer, aided by mathematical methods developed by Leonard Euler, developed tables of the moon's motions in 1757 which could be used for the determination of longitude.

At about the same time as Mayer developed his ephemeris of the moon, the British craftsman John Harrison invented the marine chronometer. Mayer's ephemeris and Harrison's chronometer were tested at sea. Observations of the moon provided longitude within four minutes of arc, and Harrison's chronometer provided longitude within one minute of arc. The British prize money was ultimately divided among Mayer's widow, Euler, and Harrison.

Thus the invention of clocks provided the means for the determination of longitude - the Earth rotates once every 24 hours, therefore the time difference between local noon and local noon at a reference meridian times the Earth's rotational speed is the angle between the observer's meridian and the reference meridian. Obviously, longitude can also be determined by observation of other astronomical bodies, given an accurate chronometer and a knowledge (an ephemeris) of the time any such body is directly overhead at some location of known longitude. Thus the accurate determination of longitude is a function of clock accuracy.

As the accuracy of measurements improved, more shortcomings in the navigation process became evident. Notably, the Earth is not a perfect sphere, the Earth has an equatorial bulge produced by centrifugal forces. Maps drawn with reference to local landmarks will not "match up to" adjacent maps of differing latitude.

Determination of position is thus dependent upon the accuracy of the clock used and the measurements of angles between the non-spherical Earth and celestial bodies. Therefore, a standard definition of the shape of the Earth and the highest accuracy possible with respect to time is necessary for the accurate determination of position.

1.1 The Reference Ellipsoid

The Earth closely resembles an oblate spheroid or ellipsoid of revolution, where the polar radius is less than the average equatorial radius, i.e., the ellipsoid of revolution

$$\frac{p^2}{a^2} + \frac{z^2}{b^2} = 1, a > b, \quad (1.1)$$

where p is the radial distance from the origin in the x, y (equatorial) plane. If one defines a reference ellipsoid that is the "best fit" to the shape of the Earth, one could then define position on the Earth's surface with respect to this reference ellipsoid. The intersection of a plane passing through the equator of the reference ellipsoid (the equatorial plane) describes a circle, and the

intersection of a plane passing through the poles of the Earth (a polar plane) describes an ellipse. The ellipses so described are called meridians. Latitude is then defined as the angle ϕ that a normal to the polar ellipse (tangent plane t) at the observer's position Q makes with the equatorial plane, see Figure 1.1. The distance from Q to the polar axis at C along the normal n is the radius of curvature in the prime vertical N at Q . Note that the normal to the ellipse does not pass through the center of the ellipse.

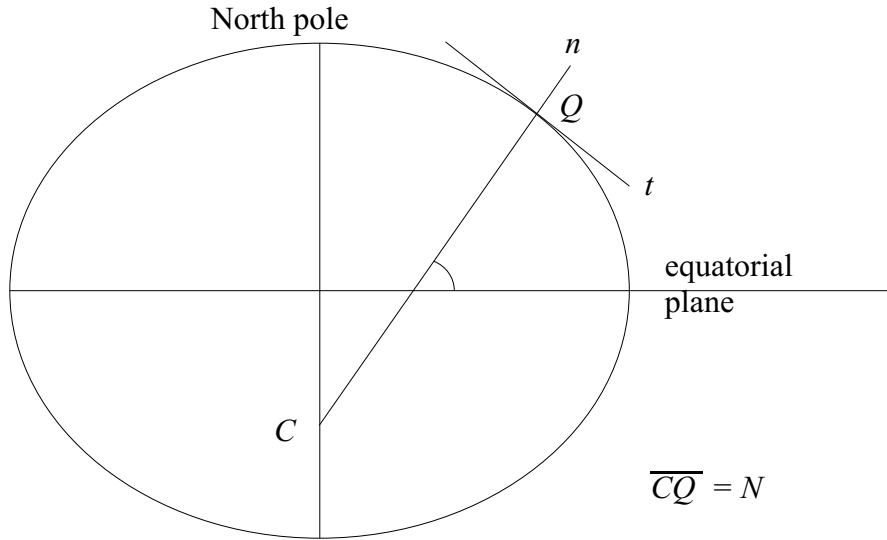


Figure 1.1: Latitude.

Longitude, λ , is defined as the angle between the polar plane passing through the reference meridian (which passes through Greenwich, England) and the polar plane containing the meridian intersecting the observer's position (see Figure 1.2).

To describe a reference ellipsoid, one needs to define, among other things, the length of the minor axis, i.e., the polar radius b (in meters), and the length of the major axis, i.e., the equatorial radius a (in meters). Typically, the reference ellipsoid is defined by the equatorial radius, a , and the non-dimensional flattening parameter, f , where

$$f = \frac{a - b}{a},$$

and the ratio of the semi axes, after re-arrangement, is

$$\frac{b}{a} = 1 - f.$$

Several reference ellipsoids have been defined over the years and are used in North America: Clarke 1866 (NAD 27), where $a = 6378206.4$ meters and

$$f = \frac{1}{294.9786982},$$

the Geodetic Reference System 1980 (NAD 83), where $a = 6378137$ meters and

$$f = \frac{1}{298.257222101},$$

the World Geodetic System 1984 (WGS 84) where $a = 6378137$ meters and

$$f = \frac{1}{298.257223563}.$$

The Global Positioning System (GPS) uses the WGS 84 ellipsoid. In addition to the flattening and the lengths of the axes, the WGS 84 ellipsoid defines the angular velocity of the Earth,

$$\dot{\Omega}_e = 7.2921151467 \times 10^{-5} \text{radians/second},$$

and the value of the Earth's universal gravitational parameter,

$$\mu = 3.986005 \times 10^{14} \text{meters}^3/\text{second}^2.$$

To specify position near the surface of the reference ellipsoid one specifies the latitude ϕ , the longitude λ , longitude radius of curvature in prime vertical N , and the height, h , above the reference ellipsoid (see Figure 1.2).

To convert from ellipsoidal coordinates ϕ, λ, h to Cartesian coordinates, first find the projection of Q (Figure 1.1) onto the equatorial plane, which is a radial distance of

$$p = N \cos \phi$$

from the polar axis, where p is the p of (1.1). The projection of that point in the equatorial plane onto the x axis is therefore

$$x = N \cos \phi \cos \lambda,$$

the projection onto the y axis is

$$y = N \cos \phi \sin \lambda.$$

We need to find N as a function of f and ϕ . First take the derivative of z with respect to p in (1.1),

$$\frac{dz}{dp} = -\frac{(1-f)^2 p}{z}. \quad (1.2)$$

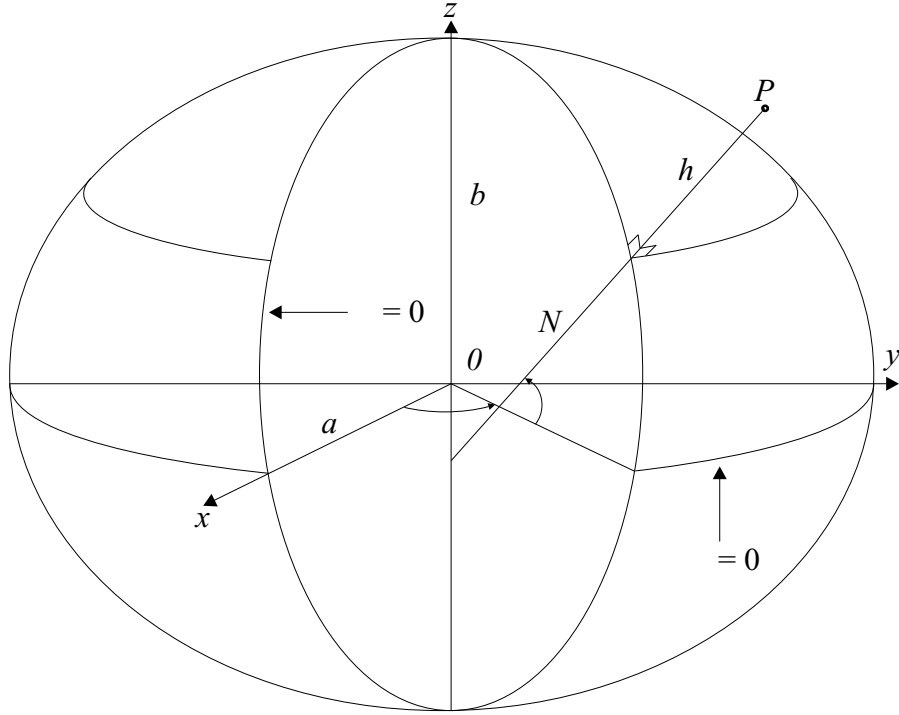


Figure 1.2: Ellipsoidal Coordinates.

Therefore the angle ϕ (in a polar plane) that the normal to the plane tangent to the reference ellipsoid makes with the equatorial plane is

$$-\frac{dp}{dz} = \frac{z}{(1-f)^2 p} = \tan \phi = \frac{\sin \phi}{\cos \phi}. \quad (1.3)$$

We are now able to express z as a function of N and ϕ :

$$\begin{aligned} z &= (1-f)^2 p \tan \phi \\ &= (1-f)^2 N \sin \phi. \end{aligned}$$

Then substituting $p = N \cos \phi$, into (1.3):

$$\frac{z}{(1-f)^2 N \cos \phi} = \frac{\sin \phi}{\cos \phi},$$

so that

$$N = \frac{z}{(1-f)^2 \sin \phi},$$

and squaring,

$$N^2 = \frac{z^2}{(1-f)^4 \sin^2 \phi}.$$

From (1.1) we obtain

$$\begin{aligned} z^2 &= b^2 - \frac{b^2 p^2}{a^2} \\ &= b^2 - (1-f)^2 N^2 \cos^2 \phi \end{aligned}$$

so that

$$N^2 = \frac{b^2 - (1-f)^2 N^2 \cos^2 \phi}{(1-f)^4 \sin^2 \phi}.$$

Then simplifying and solving for N^2 ,

$$N^2 = \frac{a^2}{(f^2 - 2f) \sin^2 \phi + 1},$$

so that

$$N = \frac{a}{\sqrt{(f^2 - 2f) \sin^2 \phi + 1}}.$$

Hence, by way of defining flattening f in terms of the parameters a and b and defining the radius of curvature, N , in terms of ϕ and the ellipsoidal parameters, the coordinates λ and ϕ of a point on the reference ellipsoid are expressed in terms of (x, y, z) :

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} N \cos \phi \cos \lambda \\ N \cos \phi \sin \lambda \\ (1-f)^2 N \sin \phi \end{bmatrix}.$$

If the height h of a point above the reference ellipsoid is specified, the radius of curvature is increased so that the coefficients of the x and y conversion become $N + h$, and z is increased by $\Delta z = h \sin \phi$:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (N + h) \cos \phi \cos \lambda \\ (N + h) \cos \phi \sin \lambda \\ ((1-f)^2 N + h) \sin \phi \end{bmatrix}.$$

If we are now given a point (x', y', z') in space above the reference ellipsoid and we want to express this point in terms of its projection (ϕ, λ, h) , onto the reference ellipsoid, we can readily determine that

$$\lambda = \begin{cases} \tan^{-1} \left(\frac{y'}{x'} \right), & \text{when } x' \neq 0 \\ \frac{\pi}{2}, & \text{when } x' = 0, \text{ and } y' > 0, \\ -\frac{\pi}{2}, & \text{when } x' = 0, \text{ and } y' < 0 \end{cases} \quad (1.4)$$

is the expression for longitude. Finding the height, h , and the latitude is not quite so straight forward.

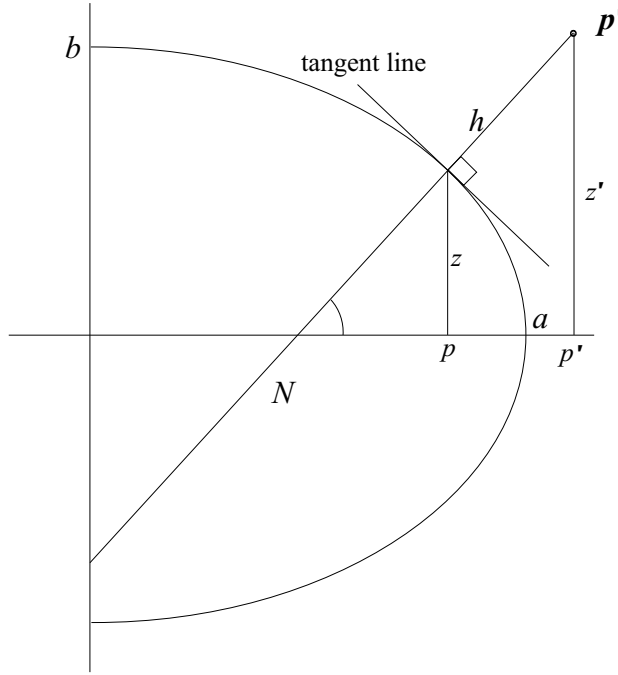


Figure 1.3: Conversion From Cartesian To Ellipsoidal Coordinates.

Let the reference ellipse (in the polar p, z plane cutting the reference ellipsoid and containing (x', y', z') and the origin) be parameterized by p , i.e.,

$$z = \frac{b}{a} \sqrt{a^2 - p^2}, \quad (1.5)$$

and the point (x', y', z') in the p, z plane is

$$\mathbf{p}' = \left(\sqrt{x'^2 + y'^2}, z' \right) = (p', z').$$

(See Figure 1.3.) Consider the reference ellipse in the first and fourth quadrants, hence $0 < p < a$ and the values of p' are by definition always positive. We further restrict our attention to the first quadrant only, fourth quadrant values of z' and z are a mirror of their first quadrant counterparts, i.e., the sign of ϕ is the same as the sign of z and z' . Hence, we consider only the absolute values of z' and choose only the positive root in (1.5) for z . Furthermore, by inspection of Figure 1.3 we note that if $p' = 0$, $\phi = 90^\circ$ and if $z' = 0$, $\phi = 0^\circ$.

A line through (p', z') normal to a tangent line at (p, z) on the reference ellipse intersects the p axis at the latitude angle ϕ . The distance from any point (p, z) on the reference ellipse to the point (p', z') is h . Then

$$h^2 = (p' - p)^2 + (z' - z)^2,$$

and substituting the parametric value for z from (1.5),

$$h^2 = z'^2 - \frac{2bz'\sqrt{a^2 - p^2}}{a} + \frac{a^2b^2 - b^2p^2}{a^2} + p'^2 - 2pp' + p^2. \quad (1.6)$$

The distance from (p', z') to a point on the reference ellipse has a minimum as p is varied about the value describing the tangent point (p, z) . We take the derivative of h^2 with respect to p in (1.6) and set it to zero to determine the value of p resulting in a point on the reference ellipse closest to (p', z') :

$$\frac{d(h^2)}{dp} = \left(\frac{2bz'}{a\sqrt{a^2 - p^2}} - \frac{2b^2}{a^2} + 2 \right) p - 2p' = 0.$$

This expression is then solved iteratively for p , and using the resulting value of p one determines h from (1.6) and the latitude ϕ from (1.3). This or another similar iteratively determined solution, or some close approximation method is commonly used to determine latitude and height above the reference ellipsoid.

Another reference figure that is used to describe the Earth is the *reference geoid*. The reference geoid is an equipotential surface, that surface that would be described if the Earth were covered with water. The reference geoid corresponds to sea level, but due to variations in the Earth's gravitational field, sea level varies from the reference geoid by several meters over the surface of the Earth. The reference geoid is not used in the GPS, but I mention it here so as to avoid any confusion between the reference geoid and the reference ellipsoid.

2 Using Artificial Satellites For Navigation

With the launch of the first artificial satellite came the possibility of using artificial satellites as the astronomical body used for the determination of position and navigation. A satellite's orbit is predictable and measurable, and if one knows the position, (x, y, z) , of three satellites relative to the Earth's center and the distance from each satellite to an observer on or near the Earth at a particular instant in time, one would be able to calculate the observer's position in relation to the center of the Earth. This is accomplished by *trilateration* – the determination of position derived from distances, as opposed to *triangulation*, where position is determined by the measurement of angles. Such a system would, however, demand that the observer's clock be perfectly synchronized with the clock that is used to determine the satellite's orbital position.

Practical considerations preclude having all observers' clocks synchronized perfectly, so a more practical system wherein the satellites provide timing signals to synchronize the observer's clock dictates that the observer receive signals from four satellites to provide the values of four unknowns: x, y, z, t , the three spatial coordinates and time. Synchronization of clocks in widely separated and inaccessible satellites is not practical, so providing highly stable clocks on the satellites and a correction factor to the master time source (which is used to determine the satellites' orbital position) is a more practical solution.

Consider then the determination of the satellite's distance from an observer. The satellite broadcasts timing information (and time correction information) to all observers, and the timing signals take a finite amount of time to reach the observer, the time determined by the the distance travelled and speed of propagation of the radio waves through the intervening medium. The result from the observer's perspective is the reception of four time signals relatively offset in time, i.e., with respect to the first signal to arrive, the other three signals will have different time delays depending upon each satellite's distance from the observer.

The first step in determining an observer's position is determining the position of the satellite as precisely as possible. The motion of a satellite relative to the Earth is governed by the law of gravitational attraction and the relationships among force, mass, and acceleration. Therefore, we shall develop the equations of motion based on these relationships.

2.1 Satellite Orbits

The first significant laws regarding the motion of satellites about a larger body are Kepler's laws. Johannes Kepler (1571-1630) had been the assistant to Tycho Brahe (1546-1601), the last great astronomer to make observations without the use of a telescope. Brahe had meticulously collected data on the motions of the planets, and Kepler spent a large part of his life studying the data and he observed regularities in planetary motion which he formulated into three laws which now bear his name:

1. The orbit of a planet describes an ellipse, with the sun at one focus.
2. The rate of description of area by the radius vector is constant, i.e., the radius vector sweeps out equal areas in equal times.
3. The cube of the semi-major axis of a planet's orbit is proportional to the square of the planet's orbital period.

Elliptical orbits are now frequently called *Keplerian* honoring Kepler's significant contribution to astronomy. Kepler derived the three laws empirically, and Sir Isaac Newton (1642-1727) later proved them analytically.

2.1.1 Newton's Laws

The motion of an artificial satellite orbiting the Earth is governed by Newton's three laws of motion [14]:

1. Every body continues in its state of rest or of uniform motion in a straight line except insofar as it is compelled to change that state by an external force.
2. The rate of change of momentum of the body is proportional to the impressed force and takes place in the direction in which the force acts.
3. To every action there is an equal and opposite reaction.

If \mathbf{r} is the position vector of a particle of mass m with respect to the origin, $\mathbf{0}$, and \mathbf{v} is its velocity and \mathbf{a} is its acceleration, then velocity is the rate of change of position with respect to time, t , or

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}.$$

Acceleration is the rate of change of velocity,

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2}.$$

The linear momentum is the product of mass and velocity, $m\mathbf{v}$, and therefore angular momentum is $m\mathbf{r} \times \mathbf{v}$. Thus from Newton's second law we have

$$\mathbf{F} = \frac{d(m\mathbf{v})}{dt} = m \frac{d\mathbf{v}}{dt} = m\mathbf{a},$$

where the constant of proportionality is unity.

The relationship which defines the magnitude of the forces involved in orbital mechanics is Newton's law of universal gravitation, one of the most far reaching scientific laws ever formulated:

“Every particle in the universe attracts every other particle in the universe with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them [14].”

Given two masses m_1, m_2 , separated by a distance r , the magnitude of the attractive force F between them is

$$F = G \frac{m_1 m_2}{r^2} \tag{2.1}$$

where G is the constant of universal gravitation.

2.1.2 The Two-Body Problem

“The quintessential problem of celestial mechanics is the two-body problem” [18]. Treatment of the two-body problem is found in most texts which concern themselves with the problem of orbit determination. Examples of such treatments may be found in [14], [18], [6], [17], [8]. The development here follows that of Roy [14].

By Newton's law of universal gravitation there is a mutual attractive force between the Earth and a satellite orbiting the Earth, and that force will, by Newton's second law of motion, act along a straight line between the respective centers of mass of the Earth and the satellite. The force of attraction \mathbf{F}_1 on body m_1 acts along the vector \mathbf{r} parallel to the line joining the center of mass of m_1 and the center of mass of m_2 , and the force of attraction \mathbf{F}_2 on mass m_2 will be directed in the opposite direction. By Newton's third law,

$$\mathbf{F}_1 = -\mathbf{F}_2.$$

The magnitude of the force is determined by Newton's law of universal gravitation, (2.1); thus we have

$$\mathbf{F}_1 = G \frac{m_1 m_2}{r^2} \frac{\mathbf{r}}{r}.$$

Newton considered, for example, the Earth and the moon, and stated and solved the two-body problem: “Given at any time the positions and velocities of two massive particles moving under their mutual gravitational force, the masses also being known, calculate their positions and velocities for any other time” [14].

The solution of the two-body problem applies to any two massive bodies, and thus forms the basis of the solution to determine the orbital parameters of an artificial satellite orbiting the Earth. Obviously, the Earth and the artificial satellite are not the only two bodies in the universe, but as a first approximation, the solution of the differential equation representing the two-body problem will yield orbital parameters to which may be applied correction factors that will take into account the other (smaller) forces acting on the artificial satellite.

Let the Earth be of mass m_1 and the satellite be of mass m_2 and let the position vector of the Earth be \mathbf{r}_1 and the position vector of the satellite be \mathbf{r}_2 with respect to some fixed reference point $\mathbf{0}$ (see Figure 2.1).

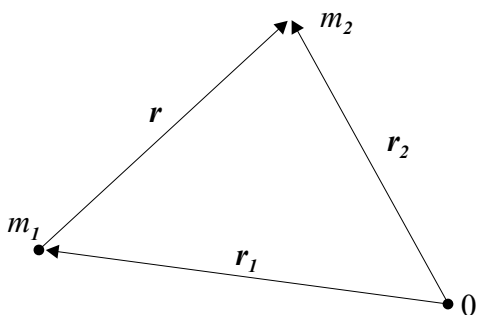


Figure 2.1: Earth And Satellite Position Vectors.

The force of attraction \mathbf{F}_1 on the Earth is directed along the vector \mathbf{r} in the direction of the line from the center of mass of the Earth to the center of mass of the satellite. (Note that the initial motion of the satellite in this analysis is not necessarily in the direction of the force applied due to the presence of the Earth. Only the initial positions and velocities of the bodies are given.)

Then

$$\mathbf{r}_1 - \mathbf{r}_2 = -\mathbf{r}.$$

Equating mass times acceleration with force, we have

$$m_1 \frac{d^2 \mathbf{r}_1}{dt^2} = G \frac{m_1 m_2}{r^2} \frac{\mathbf{r}}{r} \quad (2.2)$$

and

$$m_2 \frac{d^2 \mathbf{r}_2}{dt^2} = -G \frac{m_1 m_2}{r^2} \frac{\mathbf{r}}{r}. \quad (2.3)$$

Adding equations (2.2) and (2.3) yields

$$m_1 \frac{d^2 \mathbf{r}_1}{dt^2} + m_2 \frac{d^2 \mathbf{r}_2}{dt^2} = 0. \quad (2.4)$$

Integrating (2.4) with respect to t , we obtain the value of the momentum of the system of two massive particles, \mathbf{a}_s ,

$$m_1 \frac{d\mathbf{r}_1}{dt} + m_2 \frac{d\mathbf{r}_2}{dt} = \mathbf{a}_s$$

and integrating again with respect to t , we obtain the value of initial position of the system of two massive particles, \mathbf{b}_s ,

$$m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2 = \mathbf{a}_s t + \mathbf{b}_s$$

where \mathbf{a}_s and \mathbf{b}_s are the constant vectors of integration. Let \mathbf{R} be the position vector of the center of mass, M , of the Earth and satellite system, i.e.,

$$M \mathbf{R} = m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2$$

where $M = m_1 + m_2$. Hence

$$M \mathbf{R} = \mathbf{a}_s t + \mathbf{b}_s$$

and

$$M \frac{d\mathbf{R}}{dt} = m_1 \frac{d\mathbf{r}_1}{dt} + m_2 \frac{d\mathbf{r}_2}{dt} = \mathbf{a}_s,$$

showing that the center of mass of the Earth-satellite system moves with constant velocity, because \mathbf{a}_s is a constant vector. (Momentum and total mass are constant, therefore velocity must be constant because momentum is the product of mass and velocity.)

Dividing equations (2.2) and (2.3) by m_1 and m_2 respectively, and simplifying, we have

$$\frac{d^2 \mathbf{r}_1}{dt^2} = \frac{G m_2 \mathbf{r}}{r^3} \quad (2.5)$$

$$\frac{d^2 \mathbf{r}_2}{dt^2} = -\frac{G m_1 \mathbf{r}}{r^3}. \quad (2.6)$$

Subtracting (2.6) from (2.5) yields

$$\frac{d^2}{dt^2} (\mathbf{r}_1 - \mathbf{r}_2) = G(m_1 + m_2) \frac{\mathbf{r}}{r^3}.$$

Because $\mathbf{r}_1 - \mathbf{r}_2 = -\mathbf{r}$, it follows that

$$-\frac{d^2 \mathbf{r}}{dt^2} = G(m_1 + m_2) \frac{\mathbf{r}}{r^3}.$$

Let

$$G(m_1 + m_2) = \mu;$$

then

$$\frac{d^2 \mathbf{r}}{dt^2} + \frac{\mu \mathbf{r}}{r^3} = 0 \quad (2.7)$$

describes the motion of m_2 with respect to m_1 . Then taking the vector product of (2.7) with \mathbf{r} we have

$$\mathbf{r} \times \frac{d^2 \mathbf{r}}{dt^2} = 0.$$

Note that since

$$\frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} = 0,$$

we can add this to the previous result to obtain

$$\frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} + \mathbf{r} \times \frac{d^2 \mathbf{r}}{dt^2} = \frac{d}{dt} \left(\mathbf{r} \times \frac{d\mathbf{r}}{dt} \right) = \frac{d}{dt} (\mathbf{r} \times \mathbf{v}) = 0.$$

It therefore follows that

$$\int \left(\mathbf{r}(t) \times \frac{d^2 \mathbf{r}}{dt^2} \right) dt = \mathbf{r} \times \mathbf{v} = \mathbf{h}, \quad (2.8)$$

where \mathbf{h} is a constant vector of integration and lies in the direction of the angular momentum of the Earth/satellite system. Because the angular momentum is constant for all t , the motion of the Earth/satellite system lies in a plane orthogonal to the direction of \mathbf{h} .

Now we place the Earth (mass m_1) at the origin, and using polar coordinates, the satellite (mass m_2) is located at (\mathbf{r}, θ) , where θ is measured with respect to some arbitrary fixed line l in the plane. (See Figure 2.2).

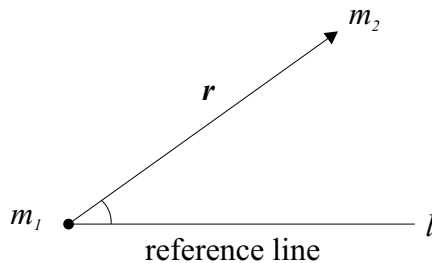


Figure 2.2: Satellite Polar Vector, Earth Origin.

Let the polar unit vector \mathbf{I} lie in the direction of \mathbf{r} , let the polar unit vector \mathbf{J} be orthogonal to \mathbf{I} , and let the polar unit vector \mathbf{K} lie in the direction

orthogonal to the plane containing \mathbf{I} and \mathbf{J} . Therefore, in terms of the Cartesian unit vectors \mathbf{i}, \mathbf{j}

$$\begin{aligned}\mathbf{I} &= \mathbf{i} \cos \theta + \mathbf{j} \sin \theta, \\ \mathbf{J} &= \mathbf{i}(-\sin \theta) + \mathbf{j} \cos \theta.\end{aligned}$$

Differentiating the polar unit vectors with respect to θ yields

$$\frac{d\mathbf{I}}{d\theta} = \mathbf{J}$$

and

$$\frac{d\mathbf{J}}{d\theta} = -\mathbf{I}.$$

The position vector of the satellite is therefore

$$\mathbf{r} = \mathbf{I}r.$$

Differentiating this product with respect to time using the chain rule, we obtain the velocity

$$\frac{d\mathbf{r}}{dt} = \frac{d}{dt}(\mathbf{I}r) = \frac{d\mathbf{I}}{d\theta} \frac{d\theta}{dt} r + \mathbf{I} \frac{dr}{dt}$$

or

$$\mathbf{v} = \mathbf{I} \frac{dr}{dt} + \mathbf{J} r \frac{d\theta}{dt}. \quad (2.9)$$

Substituting (2.9) into (2.8) yields

$$\mathbf{r} \times \mathbf{v} = \mathbf{I}r \times \left(\mathbf{I} \frac{dr}{dt} + \mathbf{J} r \frac{d\theta}{dt} \right) = r^2 \frac{d\theta}{dt} \mathbf{K} = \mathbf{h},$$

since $\mathbf{I} \times \mathbf{J} = \mathbf{K}$ and $\mathbf{I} \times \mathbf{I} = 0$ from the definition of the vector product. Therefore the magnitude of \mathbf{h} is given by

$$r^2 \frac{d\theta}{dt} = h. \quad (2.10)$$

Note the constant relationship between the length of \mathbf{r} and the time rate of change of θ . This is an expression of Kepler's second law: the vector \mathbf{r} sweeps out equal areas in equal times. Equation (2.10) is a separable differential equation. If the equation is integrated over some time period T representing 2π radians, it follows that

$$\int_0^{2\pi} r^2 d\theta = h \int_0^T dt$$

or

$$2 \int_0^{2\pi} dA = hT, \quad (2.11)$$

where dA is the polar element of area,

$$dA = \frac{r^2}{2} d\theta.$$

Hence it is evident that the constant h is twice the rate of description of area, and the value of hT is twice the area enclosed by the orbit.

Differentiating (2.9) again with respect to time using the chain rule, we obtain the polar components of acceleration:

$$\begin{aligned} \frac{d}{dt} \left(\mathbf{I} \frac{dr}{dt} + \mathbf{J} r \frac{d\theta}{dt} \right) &= \frac{d\mathbf{I}}{dt} \frac{dr}{dt} + \mathbf{I} \frac{d^2r}{dt^2} + \frac{d\mathbf{J}}{dt} r \frac{d\theta}{dt} + \mathbf{J} \frac{dr}{dt} \frac{d\theta}{dt} + \mathbf{J} r \frac{d^2\theta}{dt^2} \\ &= \frac{dr}{dt} \frac{d\mathbf{I}}{d\theta} \frac{d\theta}{dt} + \mathbf{I} \frac{d^2r}{dt^2} + \frac{d\mathbf{J}}{d\theta} \frac{d\theta}{dt} r \frac{d\theta}{dt} + \mathbf{J} \frac{dr}{dt} \frac{d\theta}{dt} + \mathbf{J} r \frac{d^2\theta}{dt^2} \\ &= \mathbf{I} \frac{d^2r}{dt^2} + 2\mathbf{J} \frac{dr}{dt} \frac{d\theta}{dt} - \mathbf{I} r \left(\frac{d\theta}{dt} \right)^2 + \mathbf{J} r \frac{d^2\theta}{dt^2} \\ &= \mathbf{I} \left[\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right] + \mathbf{J} \left[r \frac{d^2\theta}{dt^2} + 2 \frac{dr}{dt} \frac{d\theta}{dt} \right] \end{aligned}$$

so that we may express acceleration in polar terms

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{I} \left[\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right] + \mathbf{J} \left[\frac{1}{r} \frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) \right]. \quad (2.12)$$

Then substituting $\mathbf{I}r$ for \mathbf{r} into (2.7)

$$\frac{d^2\mathbf{r}}{dt^2} + \frac{\mu\mathbf{I}}{r^2} = 0,$$

and further substituting (2.12) for the acceleration term in (2.7) we obtain

$$\mathbf{I} \left[\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right] + \mathbf{J} \left[\frac{1}{r} \frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) \right] + \frac{\mu}{r^2} \mathbf{I} = 0.$$

Therefore

$$\mathbf{I} \left[\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right] + \mathbf{J} \left[\frac{1}{r} \frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) \right] + \frac{\mu}{r^2} \mathbf{I} = 0\mathbf{I} + 0\mathbf{J}.$$

Equating coefficients of \mathbf{I} and \mathbf{J} on both sides of the equation it follows that

$$\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 + \frac{\mu}{r^2} = 0,$$

or

$$\frac{d^2r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 = -\frac{\mu}{r^2} \quad (2.13)$$

and

$$\frac{1}{r} \frac{d}{dt} \left(r^2 \frac{d\theta}{dt} \right) = 0. \quad (2.14)$$

Note that multiplying (2.14) by r and integrating with respect to t yields

$$r^2 \frac{d\theta}{dt} = h$$

in agreement with (2.10).

If one eliminates time between (2.10) and (2.13) and substitutes

$$u = \frac{1}{r},$$

$$\frac{1}{r^2} = -\frac{du}{dr}$$

and

$$\frac{2}{r^3} = \frac{d^2u}{dr^2},$$

then u can be expressed in terms of θ . To accomplish this, we proceed as follows.

Because

$$\frac{dr}{dt} = \frac{dr}{d\theta} \frac{d\theta}{dt},$$

it follows that

$$\frac{d^2r}{dt^2} = \frac{d^2r}{d\theta^2} \left(\frac{d\theta}{dt} \right)^2 + \frac{dr}{d\theta} \frac{d^2\theta}{dt^2}$$

by the chain rule. From (2.10) we obtain

$$\frac{d\theta}{dt} = \frac{h}{r^2}$$

and therefore

$$\frac{d^2\theta}{dt^2} = \frac{d}{dt} \left(\frac{h}{r^2} \right) = -2 \frac{h}{r^3} \frac{dr}{d\theta} \frac{d\theta}{dt}.$$

Substituting the foregoing expressions into (2.13) we obtain

$$\frac{d^2r}{d\theta^2} \left(\frac{h}{r^2}\right)^2 - 2\frac{dr}{d\theta} \frac{h}{r^3} \frac{dr}{d\theta} \frac{h}{r^2} - r \left(\frac{h}{r^2}\right)^2 = \frac{-\mu}{r^2}.$$

After simplifying this result we have

$$\frac{d^2r}{d\theta^2} \frac{1}{r^2} - \frac{2}{r^3} \left(\frac{dr}{d\theta}\right)^2 - \frac{1}{r} = \frac{-\mu}{h^2},$$

or,

$$\frac{d^2r}{d\theta^2} \frac{du}{dr} + \frac{d^2u}{dr^2} \left(\frac{dr}{d\theta}\right)^2 + u = \frac{\mu}{h^2}.$$

Observe that

$$\frac{d^2u}{d\theta^2} = \frac{d}{d\theta} \left(\frac{du}{dr} \frac{dr}{d\theta}\right) = \frac{d^2u}{dr^2} \left(\frac{dr}{d\theta}\right)^2 + \frac{du}{dr} \frac{d^2r}{d\theta^2},$$

and therefore

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{h^2},$$

which is a linear second order differential equation. Solving for u , we have

$$u = \frac{\mu}{h^2} + c_1 e^{i\theta} + c_2 e^{-i\theta},$$

where c_1 and c_2 are constants of integration. Equivalently,

$$u = \frac{\mu}{h^2} + A \cos(\theta - \omega),$$

where A and ω are different (from c_1 and c_2) constants of integration.

Substituting

$$u = \frac{1}{r}$$

yields the relationship between the length of the Earth-satellite position vector \mathbf{r} and the angle θ it makes with the reference line previously fixed in the plane:

$$r = \frac{h^2/\mu}{1 + (Ah^2/\mu) \cos(\theta - \omega)}. \quad (2.15)$$

2.1.3 The Ellipse

Realizing that (2.15) describes a conic section, recall the definition of a conic section (see Figure 2.3): the locus of a point P that moves in the plane of a fixed point F (called a *focus*) and a fixed line d (called a *directrix*), F not on d , such

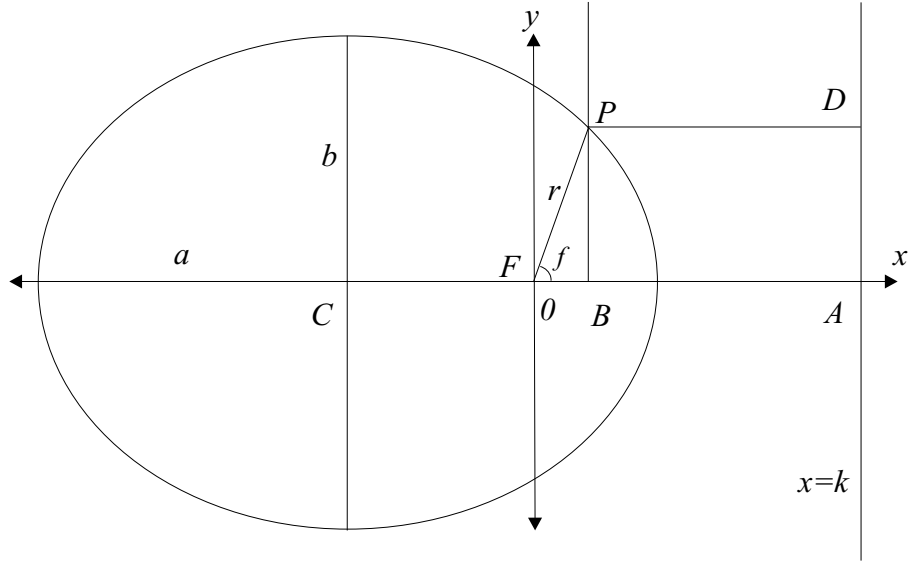


Figure 2.3: The Elliptical Locus.

that the ratio of the distance of P from F to its distance from d is a constant e (called the *eccentricity*) [3], [14].

Let the focus F be at the origin (since the Earth with mass m_1 was placed at the origin, see Figure 2.2), let the directrix d be the line $x = k$, and let the intersection of d with the x axis be the point A . Let D be the projection of the point P on d , and let B be the projection of the point P on the x axis. Let r be the line from the origin at F to P (corresponding to the satellite position vector with respect to the Earth, \mathbf{r}) and let f be the angle that r makes with the x axis (measured counterclockwise from the positive x axis). Then

$$\overline{PF} = r = (e)\overline{PD},$$

(where overlines represent distance between points) and

$$\overline{PD} = \overline{AB} = \overline{AF} + \overline{FB} = k - r \cos f$$

Then

$$r = e(k - r \cos f),$$

and, solving for r ,

$$r = \frac{ke}{1 + e \cos f}. \quad (2.16)$$

This equation is of the same form as (2.15), so now it remains to interpret (2.15) and (2.16).

In (2.16), if $e = 1$, the conic is a *parabola*; if $e < 1$, an *ellipse*; if $e > 1$, a *hyperbola*. Obviously, to be useful, an artificial satellite whose intended use is for navigation must be placed in an elliptical orbit (letting r become infinitely large is not useful!).

In the elliptical orbit described by (2.16), the point of closest approach of the satellite at P to F lies on the positive x axis; P is said to be at *perigee*. That point on the negative x axis where the ratio of that point's distance from F and its distance from d equals e is the point where P is farthest from F ; here P is at *apogee*. The center of the ellipse is midway between perigee and apogee. The distance from perigee to apogee is defined as the major diameter or major axis of the ellipse, and its length is defined as $2a$. Moving in a negative x direction from the center a distance equal to the distance from F to the center is a second focus, called the *empty focus*. The empty focus in conjunction with a second directrix located symmetrically opposite the first directrix will yield, given the same e , the same ellipse.

Consider the distances from a point P on the ellipse to the directrices. Let the distance to one directrix be α and the distance to the other directrix be β ; thus $\alpha + \beta$, the distance between the directrices, is a constant. Then the distance from a focus to P is $e\alpha$, where α is the distance from P to the directrix associated with that focus; the distance from the other focus to P is therefore $e\beta$, and $e\alpha + e\beta$ is a constant. This is another definition of an ellipse: the locus of a point whose sum of the distances to two fixed points (the foci) is a constant. Using this definition it is seen that the distance from the empty focus to the point of perigee plus the distance from perigee to the other focus is $2a$, thus the constant in this definition is the length of the major axis.

The line perpendicular to the x axis through the center of the ellipse and intersecting the ellipse in two places defines the minor axis or minor diameter which is defined to be of length $2b$. The distance from a focus to the end of the minor axis is therefore a , and thus the distance from the minor axis to a directrix is a/e . The distance from the focus to the center of an ellipse is $\sqrt{a^2 - b^2}$. Then the distance from the point of perigee to the directrix is $a/e - a$. The ratio of the distance from the focus to the point of perigee to the distance from the point of perigee to the directrix is e ; therefore the distance from the focus to the point of perigee is $a(1 - e)$. Then the distance from the focus to the center is

$$\frac{a}{e} - \left(a(1 - e) + \frac{a}{e} - a \right) = ae = \sqrt{a^2 - b^2}$$

so that

$$b = a\sqrt{1 - e^2}, \tag{2.17}$$

and the distance from the focus to the directrix is

$$k = \frac{a}{e} - ae.$$

Equation (2.16) becomes

$$r = \frac{a - ae^2}{1 + e \cos f}. \quad (2.18)$$

Note at this point if (2.17) is solved for e^2 we have

$$e^2 = \frac{a^2 - b^2}{a^2}$$

which is similar to, but not to be confused with flattening f of section 1.1. Both are measures of ellipticity.

The canonical equation for the ellipse in Cartesian coordinates, which will aid a later calculation, is the more familiar

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

an ellipse centered at the origin, semi-major axis a and semi-minor axis b .

2.2 The Elliptical Orbit

Comparing (2.18) with (2.15), we see that

$$\begin{aligned} e &= \frac{Ah^2}{\mu}, \\ A &= \frac{e}{a(1 - e^2)}, \\ f &= \theta - \omega \end{aligned}$$

and

$$\frac{h^2}{\mu} = a(1 - e^2). \quad (2.19)$$

Therefore we may express (2.15) in terms of a and e as

$$r = \frac{a(1 - e^2)}{1 + e \cos(\theta - \omega)}. \quad (2.20)$$

The constant of integration ω is interpreted as the vertex angle at the focus between the previously fixed line of reference l for θ and the positive x axis (see Figure 2.4). Note that f , the angle at the focus between the direction of perigee and r , is called the *true anomaly*. If the reference line is aligned with the x axis, i.e., the reference line contains the focus and the point of perigee and $\theta = f$, then (2.15) becomes

$$r = \frac{a(1 - e^2)}{1 + e \cos f}. \quad (2.21)$$

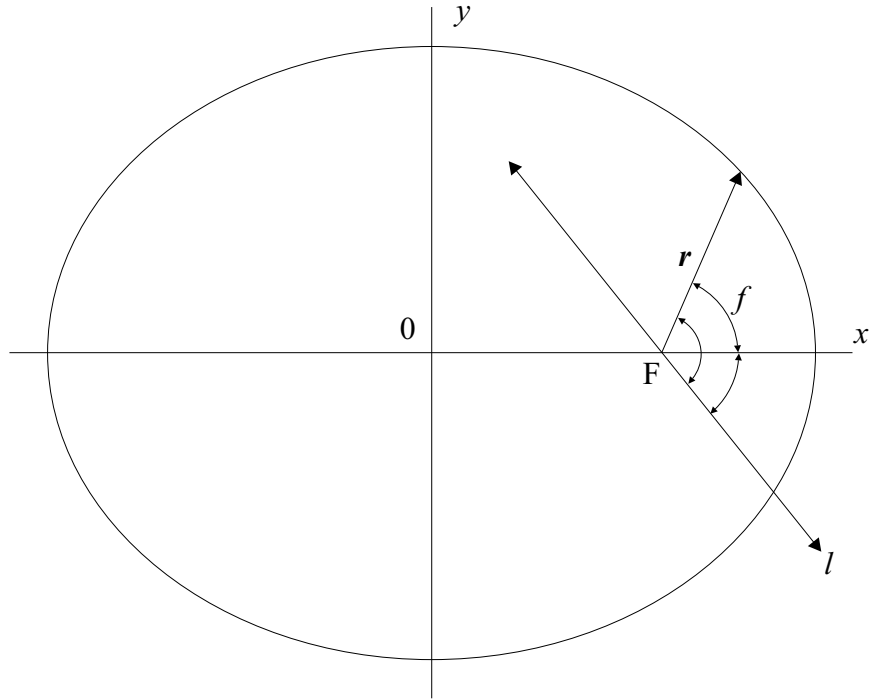


Figure 2.4: The Orbital Plane.

Because A and h^2/μ determine eccentricity, and h^2/μ is a constant, A (and hence e and a) is determined by the initial position and velocity of the satellite when it is placed into orbit. The initial position and velocity may thus be chosen not only to orient the orbit in relation to the Earth, but also to result in a highly elliptical orbit, where $e = 1 - \epsilon$, or a (nearly) circular orbit, where $e = \epsilon$, where ϵ is an arbitrarily small number. Satellites incorporated into the GPS are injected into a nearly circular orbit with a defined major axis and eccentricity so (2.20) is key to determining a satellite's position.

To use the GPS to determine one's position, one must determine the position of the satellite in its orbit from parameters that are transmitted from the satellite. In addition to major axis and eccentricity, several other aspects of orbiting bodies must be determined.

Solving (2.19) for h we obtain

$$h = \sqrt{\mu a(1 - e^2)}.$$

Because the area of an ellipse is πab , by (2.11) we also have

$$\frac{2\pi ab}{T} = h.$$

Substituting the value of b from (2.17) gives another expression for h :

$$\frac{2\pi a^2 \sqrt{1 - e^2}}{T} = h. \quad (2.22)$$

Equating the two values of h ,

$$\frac{2\pi a^2 \sqrt{1 - e^2}}{T} = \sqrt{\mu a(1 - e^2)},$$

and solving for the orbital period T , we observe that

$$T = 2\pi a \sqrt{\frac{a}{\mu}}. \quad (2.23)$$

This is a statement of Kepler's third law, relating the semi-major axis to the orbital period. Recalling that

$$\mu = G(m_1 + m_2),$$

it is evident that the orbital period is dependent only upon the semi-major axis of the orbit and the total mass of the Earth-satellite system (which is essentially that of the Earth).

To determine the velocity of the satellite in its orbit, recall the velocity equation (2.9), make the x axis our reference line, and therefore substitute f for θ :

$$\frac{d\mathbf{r}}{dt} = \mathbf{I} \frac{dr}{dt} + \mathbf{J} r \frac{df}{dt}.$$

Because the velocity vector is tangent to the elliptical orbit, it will suffice to find only the square of its magnitude. The square of the length of the velocity vector, V^2 , is given by

$$V^2 = \left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{df}{dt}\right)^2.$$

Now

$$\frac{dr}{dt} = \frac{dr}{df} \frac{df}{dt},$$

so differentiating (2.21) with respect to f , we have

$$\frac{dr}{df} = \frac{-a(1 - e^2)(-e \sin f)}{(1 + e \cos f)^2}.$$

Solving (2.10) for the time rate of change of f (recall $\theta = f$) gives

$$\frac{df}{dt} = \frac{h}{r^2},$$

and substitution for $1/r^2$ derived from (2.21) yields

$$\frac{df}{dt} = h \left(\frac{1 + e \cos f}{a(1 - e^2)} \right)^2.$$

Multiplying these two results, we have

$$\begin{aligned} \frac{dr}{df} \frac{df}{dt} &= h \left(\frac{a(1 - e^2)(e \sin f)}{(1 + e \cos f)^2} \right) \left(\frac{(1 + e \cos f)^2}{a^2(1 - e^2)^2} \right) \\ &= h \frac{e \sin f}{a(1 - e^2)}. \end{aligned}$$

Thus we obtain the first term of the expression for V^2 :

$$\left(\frac{dr}{dt} \right)^2 = h^2 \frac{e^2 \sin^2 f}{a^2(1 - e^2)^2}. \quad (2.24)$$

From (2.10) and (2.21) it follows that

$$\begin{aligned} r \frac{df}{dt} &= \frac{h}{r} \\ &= h \left(\frac{1 + e \cos f}{a(1 - e^2)} \right), \end{aligned}$$

so that

$$r^2 \left(\frac{df}{dt} \right)^2 = h^2 \frac{(1 + e \cos f)^2}{a^2(1 - e^2)^2}. \quad (2.25)$$

Adding (2.24) and (2.25) we obtain

$$\begin{aligned} V^2 &= \frac{h^2}{a^2(1 - e^2)^2} (e^2 \sin^2 f + (1 + e \cos f)^2) \\ &= \left(\frac{h}{a(1 - e^2)} \right)^2 (2 + 2e \cos f - (1 - e^2)) \end{aligned}$$

which, after re-arranging becomes

$$V^2 = \left(\frac{2h^2}{a(1 - e^2)} \right) \left(\frac{1 + e \cos f}{a(1 - e^2)} \right) - \frac{h^2}{a^2(1 - e^2)}.$$

Recognizing the right term of the product as $1/r$ and simplifying both terms using the relationship from (2.19), gives

$$V^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right). \quad (2.26)$$

Therefore, because μ is a constant, given a position r and given the desired semi-major axis a , the velocity (speed and direction tangent to the desired orbit) necessary to achieve a desired orbit is given by (2.26). This is how a nearly circular GPS satellite orbit, the orbital period, and the orbit orientation with respect to Earth are determined.

Rearranging (2.26) we obtain

$$a = \frac{\mu}{(2\mu/r) - V^2}; \quad (2.27)$$

substituting the value of a from (2.27) into (2.23) yields the period

$$T = 2\pi\mu \left(\frac{2\mu}{r} - V^2 \right)^{-\frac{3}{2}}.$$

2.2.1 Earth-Centered Earth-Fixed Coordinate System

Since a GPS satellite orbits the Earth (as in Figure 2.3), an Earth-Centered Earth-Fixed (ECEF) coordinate system is used. That is, the center of mass of the Earth is the origin, and from the satellite's perspective, the Earth is fixed in space, i.e., with reference to the stars. Six parameters define a satellite's position in the ECEF reference frame.

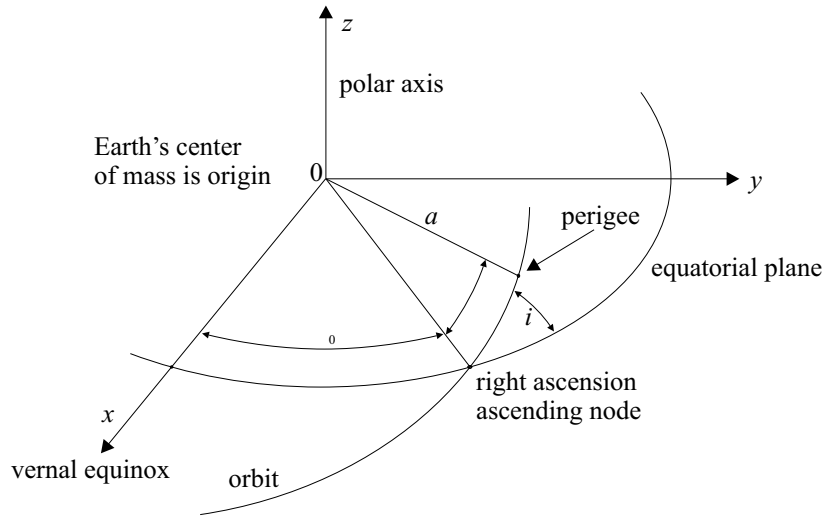


Figure 2.5: Earth-Centered, Earth-Fixed Coordinates.

We choose as our z axis the earth's rotational (polar) axis, and we shall measure longitude about this axis. The positive z axis extends from the origin

through the North pole. Thus the ECEF reference frame has a common z axis with the Earth's ellipsoidal reference frame. The intersection of the Earth's orbital plane about the sun and the Earth's equatorial plane shall be the x axis, and 0° ECEF longitude shall be in the direction defined by the positive x axis, which extends from the Earth's center of mass (the origin) and intersects the sun's center of mass at the spring, or vernal equinox. This direction is defined as the *First Point of Aries*, or the *vernal equinox*. The y axis completes a right-hand Cartesian coordinate system.

The plane containing a satellite orbit may be rotated about the x and z axes, and the ellipse describing the satellite orbit may be rotated in its plane about the origin. Therefore we define three angles (see Figure 2.5) as parameters describing a satellite orbit:

1. the *angle of inclination*, i , between the satellite's orbital plane and the Earth's equatorial plane, and
2. the angle Ω_0 , the *right ascension of the ascending node*, between the vernal equinox and the *ascending node*, where the satellite passes through the equatorial plane from south to north, and
3. the *argument of perigee*, ω , between the ascending node and the point of perigee.

(Note that the argument of perigee, ω , corresponds with ω of (2.15) and (2.20) which resulted from a constant of integration. Thus the reference line l of Figure (2.2) is here chosen to be the line intersecting the origin and the ascending node.)

The fourth and fifth parameters describing the satellite's position in the ECEF reference frame have been determined previously. Recall that one constant of integration in (2.15) determined the relationship between the semi-major axis a and the eccentricity, e . The sixth remaining parameter needed to specify a satellite's position in the ECEF reference frame is the true anomaly.

To find a satellite's position in relation to a point on the surface of the Earth, we must account for the Earth's rotation within the ECEF reference frame. (Among the parameters specified in the definition of the WGS 84 ellipsoid is the value of the Earth's angular velocity, $\dot{\Omega}_e$. See section 1.1.) The reference (Greenwich) meridian bears a relationship to the vernal equinox that is a function of time, and the angular difference between the vernal equinox and the Greenwich meridian is termed the *Greenwich hour angle*. Any meridian also bears a time relationship with the vernal equinox, since all meridians are measured from the Greenwich meridian. Thus we can relate a satellite's longitude in the ECEF reference frame (the angle between the vernal equinox and the projection of the satellite radius vector onto the equatorial plane) to longitude in the Earth's ellipsoidal reference frame.

Evident in Figure (2.5), a satellite's latitude in the Earth's ellipsoidal reference frame is a function of the angle of inclination, i , and the sum of the argument of perigee, ω , and true anomaly, f . Therefore, a satellite's position (x, y, z) in the ECEF reference frame can be converted to ϕ, λ, h in the Earth's ellipsoidal reference frame as described in section 1.1.

2.2.2 Mean and Eccentric Anomaly

The ultimate goal is to determine, as precisely as possible, the location of a GPS satellite in its orbit in order to determine a GPS user's position. The data from which a user's position is derived is based upon orbital observations made by ground stations and up-loaded to the satellite for transmission to users. Four of the six parameters describing satellite position are calculated from the data provided in the satellite transmissions with little change on the short term (> 4 hours). True anomaly and the radius vector \mathbf{r} change rapidly and these values must be calculated by the GPS user continuously. Note also that the true anomaly is measured at the elliptical orbit focus, while Earth-based users are referenced to the Earth's center of mass. Mean anomaly and eccentric anomaly provide the means to calculate the true anomaly and hence \mathbf{r} .

During the orbital period T , the radius vector turns through 2π radians; therefore the mean angular velocity or mean motion, n , is defined as

$$n = \frac{2\pi}{T}, \quad (2.28)$$

and using the value of T from (2.23) yields

$$n = \sqrt{\frac{\mu}{a^3}}. \quad (2.29)$$

If the radius vector of an elliptical orbit is at perigee at time τ , then at some later time t , *another* radius vector rotating at mean angular velocity n will have swept through an angle M , defined as the *mean anomaly*, where

$$M = n(t - \tau). \quad (2.30)$$

In Figure 2.6 a circle of radius a is circumscribed about an ellipse of major axis $2a$ and minor axis $2b$. The circle's radius vector, \overline{CQ} , of length a , moves such that the point of its intersection with the circle and the point of intersection of the radius vector \mathbf{r} with the ellipse lie on a perpendicular drawn to the major axis of the ellipse (the x axis). Hence the motion of the circle's radius vector is locked to the motion of the elliptical radius vector. The angle that the circle's radius vector \overline{CQ} makes with the x axis, $\angle QCR = E$, is defined as the *eccentric anomaly* and is therefore a function of true anomaly, f .

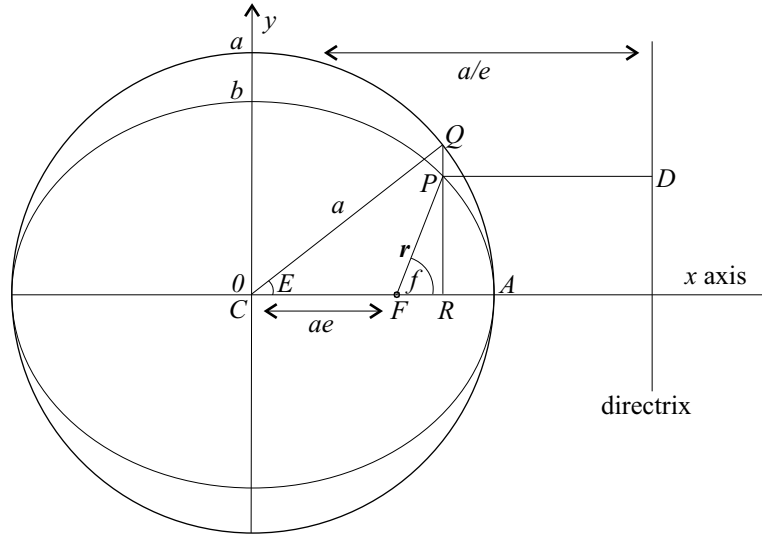


Figure 2.6: Eccentric Anomaly.

From Figure 2.6 the length of line \overline{PD} is determined by the eccentricity, and is a function of f . Because the distance from the origin to the directrix is a/e and

$$\overline{CR} = a \cos E$$

it follows that

$$\overline{PD} = \frac{a}{e} - a \cos E.$$

Using the definition of the ellipse,

$$\overline{PD} = \frac{r}{e},$$

and equating these two expressions we find that

$$r = a(1 - e \cos E). \quad (2.31)$$

Equating again the previously determined two values of \overline{PD} , but substituting the polar expression for r from (2.21), we find that

$$\frac{a}{e} - a \cos E = \frac{1}{e} \left(\frac{a(1 - e^2)}{1 - e \cos f} \right).$$

This expression may be solved for $\cos f$ or $\cos E$:

$$\cos f = \frac{e - e \cos E}{e \cos E - 1} \quad (2.32)$$

$$\cos E = \frac{\cos f + e}{1 + e \cos f}. \quad (2.33)$$

If the eccentric anomaly is known, the true anomaly may be calculated from (2.32). From Figure 2.6 observe that

$$\overline{FR} = r \cos f, \quad \overline{CR} = a \cos E, \quad \overline{FR} = a \cos E - ae;$$

hence

$$r \cos f = a \cos E - ae. \quad (2.34)$$

Further,

$$\overline{QR} = a \sin E.$$

Observe that

$$\overline{PR} = r \sin f$$

then solve the canonical equation for an ellipse for y ,

$$y^2 = b^2 \left(1 - \frac{x^2}{a^2} \right),$$

and substitute the value of \overline{CR} for x to yield

$$\begin{aligned} \overline{PR}^2 &= b^2 \left(1 - \frac{a^2 \cos^2 E}{a^2} \right) \\ &= b^2 (1 - \cos^2 E). \end{aligned}$$

Hence

$$\overline{PR} = b \sin E,$$

and therefore

$$r \sin f = b \sin E. \quad (2.35)$$

Comparing \overline{PR} with \overline{QR} we obtain

$$\frac{\overline{PR}}{\overline{QR}} = \frac{b \sin E}{a \sin E} = \frac{b}{a} \quad (2.36)$$

which is a constant. Recalling Kepler's second law (2.11) and using Figure 2.6, we see that

$$\frac{\text{area } FPA}{\text{area of ellipse}} = \frac{t - \tau}{T},$$

and since the area of an ellipse is πab ,

$$\text{area } FPA = \frac{\pi ab(t - \tau)}{T}.$$

Using (2.28) for mean motion and (2.30) for mean anomaly, it follows that

$$\text{area } FPA = \frac{abM}{2}. \quad (2.37)$$

From Figure 2.6 we see that

$$\text{area } FPA = \text{area } FPR + \text{area } RPA.$$

Observe that integrating to find the area under the arc QA and the area under the arc PA leads to

$$\text{area } RPA = \frac{b}{a} (\text{area } QRA),$$

since the ratio of the areas is defined by the relationship derived in (2.36). Then

$$\begin{aligned} \text{area } FPA &= \text{area } FPR + \frac{b}{a}(\text{area } QRA) \\ &= \text{area } FPR + \frac{b}{a}(\text{area } QCA - \text{area } QCR) \\ &= \frac{(r \sin f)(r \cos f)}{2} + \frac{b}{a} \left(\frac{a^2 E - a^2 \sin E \cos E}{2} \right). \end{aligned}$$

Substituting identities from (2.34) and (2.35) above and simplifying yields

$$\text{area } FPA = \frac{ab(E - e \sin E)}{2}.$$

Comparing this result with (2.37), we obtain Kepler's equation

$$E - e \sin E = M = n(t - \tau), \quad (2.38)$$

relating eccentric anomaly to mean anomaly, and hence mean motion. From Kepler's equation and equation (2.29) for mean motion we derive the relationship

$$E - e \sin E = (t - \tau) \sqrt{\frac{\mu}{a^3}}. \quad (2.39)$$

Then given the elapsed time since a satellite's point of perigee, the constant μ , the semi-major axis a , and the eccentricity e , equation (2.39) may be solved for E (by iteration). Then the true anomaly, f , is obtained from (2.32) and r is then determined by (2.31). GPS orbits have very low eccentricity, so iterative solutions for (2.39) converge quickly with a starting value $E = M$.

2.2.3 Perturbed Orbits

As stated earlier, the solution to the two-body problem serves as a first approximation to the actual orbit of a satellite. Other forces act upon a satellite in its orbit, such as gravitational attraction of the moon, the sun, and other planets. Solar radiation – or lack of it when the satellite is on the dark side of the Earth – exerts a force on a satellite, and collisions with other particles, e.g., atmospheric drag, also contribute forces that perturb the satellite’s orbit. The Earth itself in the two body problem is treated as a point mass, when in reality the shape of the Earth is not spherical, the mass distribution of the Earth is not homogeneous, and the Earth’s surface varies from that of the reference geoid; i.e., there are variations in the Earth’s gravitational field. “In the artificial satellite case the main perturbing effects are due to the nonspherical components of the Earth’s gravitational field and to drag by the Earth’s atmosphere” [14]. The orbits of the satellites in the GPS are high enough that atmospheric drag is not a large factor, but the Earth’s gravitational field is a perturbing factor that must be addressed in the short term. Attempts have been made to model the many forces acting on a satellite in Earth orbit, but it has been concluded that “even the two-body case, where one of the bodies is of arbitrary shape and mass distribution, cannot in general be solved in closed form” [14].

What can be done, however, is to modify the predicted two body problem solution to account for the asymmetrical shape and mass distribution of the Earth. One can calculate the coefficients of the modified equation for a best fit based on actual measurements of the satellite’s orbit. This is the approach taken in the GPS; the data message received by users of the GPS from each satellite contains frequently updated (i.e. several times a day) orbital correction coefficients to modify the Keplerian orbit prediction.

“In 1959 it was observed that the orbit of Vanguard, this country’s first satellite, followed an orbit that did not agree exactly with that calculated using values of [acceleration due to gravity] based on a near-ellipsoidal geoid. It was concluded that the geoid is best approximated, not by an ellipsoid of revolution, but by a slightly pear-shaped figure, the small end of the “pear” being in the northern hemisphere and extending about 15 m above the reference ellipsoid” [7].

Since Vanguard, careful observations of satellites’ orbits have been used to map the Earth’s gravitational field. What follows here is a brief overview of some of the findings to lend insight into the orbital calculations used for the satellites in the GPS.

The acceleration imposed on a particle located at point P a distance r from the center of mass of a planet of mass M (see Figure 2.7) is derived from (2.3)

$$\frac{d^2 \mathbf{r}}{dt^2} = -G \frac{M \mathbf{r}}{r^2 r}.$$

Let the *potential* (sometimes called the *Newtonian potential*) at P due to the presence of mass M be defined as U , where

$$U = \frac{GM \mathbf{r}}{r r}.$$

Then

$$\frac{d^2 \mathbf{r}}{dt^2} = \nabla U.$$

Gravitational acceleration imposed on a particle can therefore be derived from the potential. We shall find it easier in what follows to deal with the potential rather than acceleration.

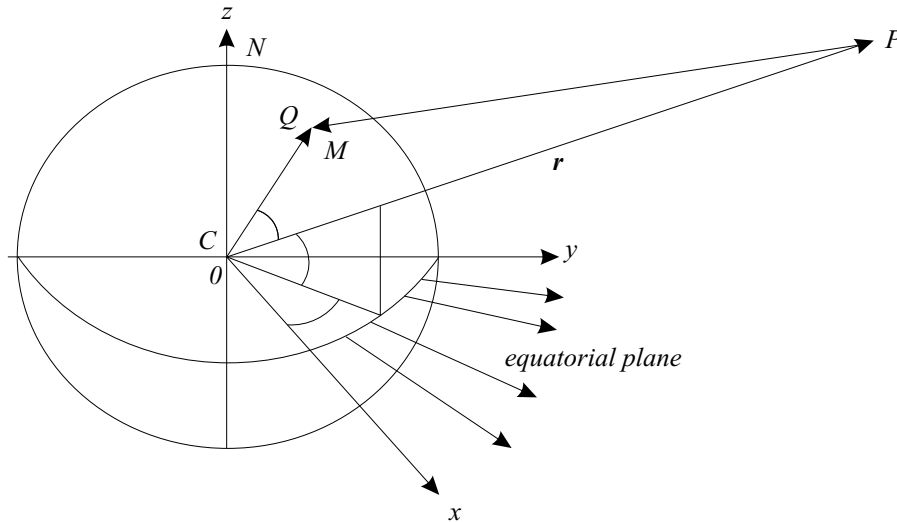


Figure 2.7: Potential Referenced To A Planet's Center Of Mass.

Then a particle of mass ΔM within the planet located at point Q a distance ρ from the center of mass contributes an increment of potential

$$\Delta U = G \frac{\Delta M}{PQ}$$

to the total potential at P due to the planet. The total potential at P due to the planet is therefore

$$U = G \int_E \frac{dM}{PQ},$$

where the integral is taken over the whole body of the planet E .

Take the center of mass of the planet as the coordinate origin, and denote the coordinates of P and Q by (x, y, z) and (ξ, η, ζ) respectively. Designate the angle between \mathbf{r} and $\boldsymbol{\rho}$, i.e., $\angle P0Q$, as ψ . Then

$$\begin{aligned}\overline{PQ}^2 &= (x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2, \\ r^2 &= x^2 + y^2 + z^2,\end{aligned}\tag{2.40}$$

and

$$\rho^2 = \xi^2 + \eta^2 + \zeta^2.$$

Rearranging (2.40) and factoring out r^2 yields

$$\overline{PQ}^2 = r^2 \left[1 - 2 \left(\frac{x\xi + y\eta + z\zeta}{r\rho} \right) \frac{\rho}{r} + \left(\frac{\rho}{r} \right)^2 \right].$$

From the Law of Cosines, we have

$$\cos \psi = \left(\frac{x\xi + y\eta + z\zeta}{r\rho} \right).$$

We now let

$$\alpha = \frac{\rho}{r} \text{ and } q = \cos \psi,$$

so that

$$\overline{PQ}^2 = r^2(1 - 2q\alpha + \alpha^2).$$

Hence

$$\frac{1}{\overline{PQ}} = \frac{1}{r\sqrt{1 - 2q\alpha + \alpha^2}},\tag{2.41}$$

and

$$U = G \int_E \frac{dM}{r\sqrt{1 - 2q\alpha + \alpha^2}}.$$

The function

$$f = \frac{1}{\sqrt{1 - 2q\alpha + \alpha^2}}$$

occurs often in physics, for example finding the potential at some distance from an electrically charged particle. The function f is a *generating function* of the Legendre polynomials, $P_n(q)$. Briefly, by taking the partial derivatives of f with respect to q and α and equating coefficients of the same powers of α one can determine a recursion relation that generates the Legendre polynomials.

Therefore

$$U = \frac{G}{r} \left(\int_E P_0 dM + \int_E P_1 \alpha dM + \int_E P_2 \alpha^2 dM + \dots + \int_E P_n \alpha^n dM + \dots \right),\tag{2.42}$$

where the first five polynomials are

$$\begin{aligned} P_0 &= 1, \\ P_1 &= q, \\ P_2 &= \frac{1}{2}(3q^2 - 1), \\ P_3 &= \frac{1}{2}(5q^3 - 3q), \\ P_4 &= \frac{1}{8}(35q^4 - 30q^2 + 3). \end{aligned}$$

Then (2.42) may be written

$$U = \sum_{i=0}^{\infty} U_i$$

where

$$U_i = \frac{G}{r} \int_E P_i \alpha^i dM, \quad i = 0, 1, 2, \dots$$

For the purpose of finding GPS satellite positions, we will be concerned with only the first three terms, $i = 0, 1, 2$. Evaluating the first integral,

$$U_0 = \frac{G}{r} \int_E dM = \frac{GM}{r},$$

the value of the mass potential with M taken as a point mass. The second term is

$$\begin{aligned} U_1 &= \frac{G}{r} \int_E q \alpha dM, \\ &= \frac{G}{r^2} \int_E \left(\frac{x\xi + y\eta + z\zeta}{r} \right) dM, \\ &= \frac{G}{r^3} \left(x \int_E \xi dM + y \int_E \eta dM + z \int_E \zeta dM \right). \end{aligned}$$

If the center of mass of a body B is chosen as the origin of a Cartesian coordinate reference frame (x, y, z) , then

$$\int_B x dM = \int_B y dM = \int_B z dM = 0.$$

Since we have chosen the center of mass of the planet E as our origin,

$$\int_E \xi dM = \int_E \eta dM = \int_E \zeta dM = 0,$$

therefore it follows that

$$U_1 = 0.$$

Now consider the third term

$$U_2 = \frac{G}{2r^3} \int_E (3q^2 - 1)\rho^2 dM.$$

Substituting for q ,

$$U_2 = \frac{G}{2r^3} \left(3 \int_E (\rho \cos \psi)^2 dM - \int_E \rho^2 dM \right). \quad (2.43)$$

Note that in this equation, $\rho \cos \psi$ is the projection of ρ onto the line OP . Let

$$\rho \cos \psi = X,$$

and with X form a right-hand set of coordinate axes so that projecting ρ onto these coordinate axes yields

$$\rho^2 = X^2 + Y^2 + Z^2,$$

where Y, Z are the projections of ρ onto the other two axes. Then by substitution, (2.43) becomes

$$U_2 = \frac{G}{2r^3} \left(3 \int_E X^2 dM - \int_E (X^2 + Y^2 + Z^2) dM \right), \quad (2.44)$$

and thus

$$U_2 = \frac{G}{2r^3} \int_E (2\rho^2 - 3(Y^2 + Z^2)) dM.$$

The moment of inertia about the x axis, y axis, and the z axis respectively are defined as

$$A = \int_E (y^2 + z^2) dM,$$

$$B = \int_E (x^2 + z^2) dM,$$

$$C = \int_E (x^2 + y^2) dM,$$

where the integral is taken over the whole body, and (x, y, z) in this definition are the coordinates of the element of mass, dM . The moment of inertia about the axis defined by the line OP is therefore

$$I = \int_E (Y^2 + Z^2) dM.$$

By substitution (2.44) becomes

$$U_2 = \frac{G}{2r^3}(A + B + C - 3I), \quad (2.45)$$

and U_2 is a function of the moments of inertia. This equation is called MacCullagh's formula. If the body is spherical with a homogeneous distribution of mass, the moments about any axis are equal and we see that $U_2 = 0$ and U in this case would be

$$U = \frac{GM}{r},$$

the value obtained if the mass were considered a point.

As stated in the quotation above, the Earth's shape deviates from the reference ellipsoid, and is in reality "pear shaped;" thus $U_2 \neq 0$ in the case of the Earth. The Earth is reasonably symmetrical about any polar plane, but is asymmetrical about the equatorial plane. Thus the equatorial moment of inertia A is very nearly equal to the equatorial moment of inertia B , but A and B differ from the polar moment of inertia C .

Note that the moment of inertia about an arbitrary axis is defined to be kMg^2 kilogram-meters², where M is the mass of the body, g is a length which we shall call (for the lack of a better term) the *effective radius* of the body in the plane normal to the arbitrary axis, and k is a constant depending upon the choice of the axis. Denote the moment of inertia divided by the total mass, i.e. kg^2 as γ_n , where $n = 0, 1, 2, \dots$ and γ_n therefore denotes the particular value obtained about axis n .

An oblate spheroid with non-uniform mass distribution has different moments of inertia about the equatorial and polar axes. What is common in every case is the total mass and its distribution, and also the transition from the potential at a point in the equatorial plane to a point on the polar axis opposite either pole (the distance from the center of mass being held constant). Then the potential U_2 defined by (2.45) can be re-written factoring out the mass M from the moments of inertia:

$$U_2 = \frac{GM}{2r^3}(A' + B' + C' - 3I'). \quad (2.46)$$

Then with respect to an arbitrary axis defined by the line OP , denote

$$\gamma_0 = (A' + B' + C' - 3I').$$

Assume that A is equal to B , and define the potential, when the axis under consideration lies in the equatorial plane, i.e. $3I' = 3A'$, as U_{2A} . When the axis

under consideration is coincident with the polar axis, i.e. $3I' = 3C'$, define the potential as U_{2C} . Then

$$\begin{aligned} U_{2A} &= \frac{GM}{2r^3}(2A' + C' - 3A'), \\ &= \frac{GM}{2r^3}(C' - A'), \\ U_{2C} &= \frac{GM}{2r^3}(2A' + C' - 3C'), \\ &= \frac{GM}{2r^3}(2A' - 2C'). \end{aligned}$$

Then with respect to the moment of inertia as taken about an axis in the equatorial plane:

$$\gamma_1 = (C' - A').$$

And with respect to the moment of inertia as taken about an axis coincident with the polar axis:

$$\gamma_2 = (2A' - 2C').$$

The transition of the value of the moment of inertia from the equatorial plane to the polar axis is therefore a function of latitude. Assume the transition between the quantity γ_1 and the quantity γ_2 describes an ellipse in the first quadrant (because the Earth cut by a polar plane describes an ellipse) and therefore γ_0 as a function of latitude is

$$\gamma_0 = (C' - A') \cos^2 \phi + (2A' - 2C') \sin^2 \phi, \quad (2.47)$$

where $0 \leq \phi \leq \pi/2$ is the angle of latitude. Simplifying, we find

$$\begin{aligned} \gamma_0 &= (C' - A')(\cos^2 \phi - 2 \sin^2 \phi) \\ &= (C' - A')(1 - \sin^2 \phi - 2 \sin^2 \phi) \\ &= (C' - A')(1 - 3 \sin^2 \phi) \end{aligned}$$

Substituting γ_0 into (2.46)

$$\begin{aligned} U_2 &= \frac{GM}{2r^3}(C' - A')(1 - 3 \sin^2 \phi) \\ &= \frac{G}{2r^3}(C - A)(1 - 3 \sin^2 \phi). \end{aligned}$$

Because we assumed that A is equal to B , we do not take into account any effect due to the longitude, λ . (There is a small effect, but it will not affect our GPS calculations. The calculation of the effect of changing longitude is carried out in an analogous fashion to that used here for determining the effects of latitude.)

Based on the assumptions made, we expect that

$$U_2 = \frac{GM}{r^3}(C - A)kR^2(1 - 3\sin^2\phi),$$

where k is as yet an undetermined constant, and R is the Earth's *equatorial radius*. In relation to the Earth's equatorial radius R , U_2 can in fact be approximated by

$$U_2 = \frac{GM}{r^3}(C - A)R^2\left(\frac{1}{3} - \sin^2\phi\right),$$

where we see that $k = 1/3$, confirming the assumptions.

The accepted expression representing the Earth's gravitational potential has been determined to be

$$U = \frac{GM}{r} \left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R}{r}\right)^n P_n(\sin\phi) \right]$$

where J_2 is the coefficient of the second harmonic of the Earth's gravitational potential and "[t]he physical interpretation of J_2 is that it is the difference between the polar and equatorial moments of inertia per unit mass" [18]. The first few coefficients and constants have been determined [14] to be

$$\begin{aligned} 10^6 J_2 &= 1082.63 \pm 0.01, & 10^6 J_3 &= -2.51 \pm 0.01, & 10^6 J_4 &= -1.60 \pm 0.01, \\ 10^6 J_5 &= -0.13 \pm 0.01, & GM &= 398603.2 \text{ km}^3 \text{ s}^{-2}, & R &= 6378.165 \text{ km}. \end{aligned}$$

Thus we see that J_2 is several orders of magnitude greater than the other J_n .

The acceleration due to gravity varies not only with distance, but also latitude. Therefore, any orbit about an oblate spheroid that does not lie in an equatorial plane is affected by a changing force. This variation occurs twice (for two equatorial crossings and two approaches toward the poles) in each complete circuit around the planet and therefore affects the orbital parameters a , e , ω , Ω , and i to some extent. Any compensation for this effect has a period that is half the orbital period, i.e., is a second harmonic; note that (2.47) is such a correction. For values of any sign, say v_1 and v_2 , we similarly find the corrected value v_c :

$$v_c = v_1 \sin 2\phi + v_2 \cos 2\phi. \quad (2.48)$$

We shall use an expression of this form to correct the GPS satellite orbital parameters for the second harmonic perturbation.

3 The Global Positioning System

“The Global Positioning System (GPS) is a space-based radionavigation system managed and operated by the United States (U.S.) Government. GPS was designed as a dual-use system with the primary purpose of enhancing the effectiveness of U.S. and allied military forces. . . GPS is also becoming an integral component of the Global Information Infrastructure, with applications ranging from mapping and surveying to international air traffic management and global climate change research” [2].

The performance specifications of the GPS are elaborated in the *Global Positioning System Standard Positioning Service Performance Standard* and a few highlights from that document are presented here.

The Government divides the GPS into three segments:

1. the Space Segment,
2. the Control Segment, and
3. the User Segment.

A constellation of 24 operational satellites (a GPS satellite is referred to as a space vehicle, “SV”) constitutes the Space Segment. The design criteria call for positioning four SVs each in six equally spaced orbital planes (60° of longitude apart), nominally inclined at 0.30 semi-circles¹ (i.e., $54^\circ \pm 1.000^\circ$ with respect to the equatorial plane. The longitude of the ascending node, Ω_0 , of the first plane is nominally 42.374° . The design calls for a nominal eccentricity of 0.000 and a semi-major axis of 26559.7 kilometers. “The GPS constellation is a dynamic entity. . .” [2] so the actual values attained vary somewhat from the design parameters. Tolerances and operational ranges are specified to be:

- Groundtrack Equatorial Crossing: $\pm 2^\circ$
- Eccentricity: 0.00-0.02
- Inclination: $\pm 3^\circ$
- Semi-major Axis: ± 50 kilometers for Block IIR, ± 17 kilometers for Block II/IIIA ²

¹A semi-circle is a unit of angular measure used in the Government publications cited.

²Block IIR, Block II/IIIA are different types of GPS satellites.

- Longitude of the Ascending Node: $\pm 2^\circ$
- Argument of Perigee: $\pm 180^\circ$

Actual values of eccentricity and inclination, as of 9/24/2000, range from $e = 0.0013$ to $e = 0.0166$, and $i = 56.505^\circ$ to $i = 52.957^\circ$ [2].

From the nominal values we can calculate the nominal SV speed from (2.26) at 3873.979667 meters/second, or 13946.32680 kilometers/hour, and from (2.23) we find the nominal period is 43077.024 seconds, or 11.965840 hours.

The Control Segment consists of four major components:

1. a Master Control Station (MCS) located at Schriever Air Force Base, Colorado;
2. a Backup Master Control Station, located at Gaithersburg, Maryland;
3. four ground antennas that provide near real-time telemetry, tracking, and commanding interface among the GPS satellites and the MCS; and
4. six monitor stations that provide real-time satellite ranging measurement data to the MCS and support near continuous monitoring of the constellation performance [2].

The Control Segment monitors the performance of the GPS and provides data the User Segment needs to make use of the GPS. Data gathered on orbital parameters and timing are uploaded to each SV, and each SV then retransmits, among other things, the orbital parameters to the User Segment (those who make use of the GPS).

3.1 The Space Segment/User Segment Interface

The description of the Space Segment/User Segment Interface is contained in *Navstar GPS Space Segment/Navigation User Interfaces*, *ICD-GPS-200*, *Global Positioning System Standard Positioning Service Performance Standard* and *Global Positioning System Standard Positioning Service Signal Specification*. The information in this section is taken from these documents.

3.1.1 GPS Satellite Signals

The time and frequency standard on the SVs is a nominal 10.23 MHz source. There is a relativistic effect on the SV time/frequency source to be accounted for due to the relative velocity between the SV and the user, and the influence of the Earth's gravitational field. The SV frequency standard, nominally 10.23000000000 MHz, is set to 10.22999999543 MHz prior to launch. Thus to an

observer on the SV, the clock on the SV appears to run slow, but to the observer on Earth, the clock is on time. All SVs transmit their data on the same two L band frequencies, L1=1575.42 MHz and L2=1227.6 MHz. The carrier frequencies (integer multiples of the time/frequency source) and the data rates are derived from the time/frequency source on board the SV. The Standard Positioning Service (“SPS”) utilizes the L1 frequency, and the Precise Positioning Service (“PPS”) utilizes both L1 and L2. Data on the L2 carrier are encrypted as are the the data on L1 that pertain to the PPS. Civilian users have access to the SPS.

The radio frequency output of the SV (both carriers) is a code division multiplex modulated carrier. The modulation contains a P(Y) code, a Coarse Acquisition (“C/A”) code, and a Navigation (“NAV”) message. Each SV’s unique identification number allows the user to decode a particular SV’s data by finding the identification number code sequence in the P(Y) and C/A codes.³

The master clock (“GPS time”) on board the SV is a continuous count of 1.5 second epochs beginning with the X1 epoch beginning at midnight, January 5, 1980/morning January 6, 1980, Universal Coordinated Time (U.S.Naval Observatory) (“UTC (USNO)”). A week consists of 604,800 seconds (403,200 1.5 second epochs), and a continuously incremented week number is generated every 403,200 X1 epochs to be part of the Z count (binary) contained in the NAV message. The week number is the ten most significant bits of the twenty-nine bit Z count, and is the week number, modulo 1024, since the beginning of GPS time. The nineteen least significant bits, the Time of Week (“TOW”), of the Z count are the binary representation of the number of X1 epochs elapsed in the current week. The TOW is reset to zero at the end of every week (403,199 X1 epochs) and the week number is incremented on the first X1 epoch of the week. Note that UTC is corrected from time to time by the addition of “leap seconds” and therefore there is some (known) difference between UTC time and GPS time. “There also is an inherent but bounded drift rate between the UTC and the GPS time scales. The [Control Segment] shall control the GPS time scale to be within one microsecond of UTC (modulo one second)” [1]. The Control Segment also provides (in the form of coefficients for a time correction polynomial) the data in the NAV message to relate GPS time to UTC within 90 nanoseconds.

There is an *equipment group delay*,

. . . defined as the delay between the L-band radiated output of a specific SV (measured at the antenna phase center) and the output of that SV’s on-board frequency source; the delay consists of a bias term and an uncertainty. The bias term is of no concern to the [User]

³This is an oversimplification. Detail is provided in the Government publications cited.

since it is included in the clock correction parameters relayed in the NAV data, and is therefore accounted for by the user computations of system time. . . The effective uncertainty of the group delay shall not exceed 3.0 nanoseconds. . . [1].

There also exists a group delay differential, termed T_{GD} , between the radiated L1 and L2 P(Y) signals. Since the clock correction parameters are related to the L2 carrier signals, the SPS user must subtract this value after the application of the clock correction parameters.

Delays are incurred by the carrier frequency signal and its modulation as the signal passes through the ionosphere. Signal delay and refraction along the path between the SV and the user is a function of the electron density in the ionosphere, in electrons per meter³, along the path. The electron density in the ionosphere also varies with time of day and the solar cycle (sunspot activity), among other things. Analysis shows that in passing through the ionosphere, “. . . GPS code measurements are delayed and the carrier phases are advanced” [8]. Ionospheric delays, in meters at a frequency of 2 GHz, range from 0.1 meter at a total electron count of 10^{16} electrons/meter³ to 10 meters at a total electron count of 10^{18} electrons/meter³ [12]. Delays and refraction of the radio frequency signal also occur in passage through the troposphere, and vary by the relative amounts of dry air and water vapor along the path between the SV and the user. Accordingly, ionospheric and tropospheric delays must be modelled (and by the nature of the processes being modelled, imperfectly), and several models are discussed in the literature; the model of choice depends on the user’s preference. Descriptions of several models and their derivations are contained in *GPS Theory and Practice* and a tropospheric model is shown in *GPS Satellite Surveying*. Because electron density and water vapor/dry air ratios are a function of the angle the path between the user and SV makes with the zenith, both ionospheric and tropospheric models must address path *obliquity* (direction of path compared with zenith). A particular ionospheric delay model is prescribed in *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200* for use in the SV range calculation. That model will be shown in this thesis without elaboration, as the Government offers no explanation of its derivation and the specifications of the GPS are based on its use. Note that the ionospheric and tropospheric delays, since the models are imperfect, contribute to the error of the range calculation (and are, in fact the largest contributor to the error).

3.1.2 The NAV Message

The NAV message from the SV is composed of twenty-five 1500 bit frames transmitted in a serial bit stream at the bit rate of 50 bits per second (20 ms per

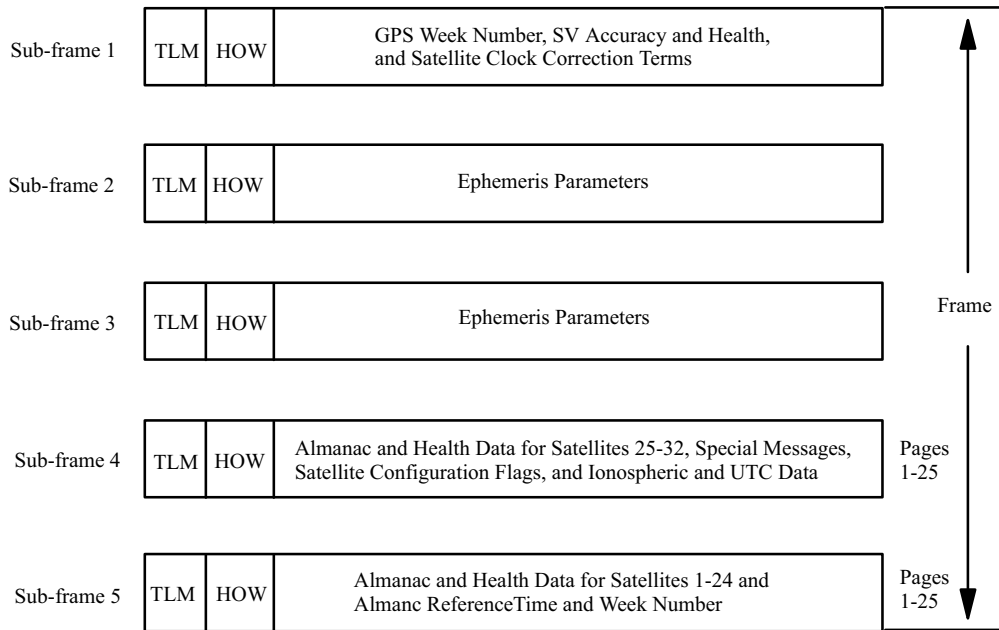


Figure 3.1: Significant Subframe Contents. (Adapted from *Global Positioning System Standard Positioning Service Signal Specification*.)

bit). Each frame is made up of five subframes, each subframe is 300 bits in length (see depiction of subframe contents, Figure 3.1) and each subframe contains 10 data words 30 bits long. Subframes number 4 and number 5 are sub-commutated 25 times, and each sub-commutation is termed a page. Therefore, a complete data message is 25 frames in length, where subframes 1 through 3 (see Figures B.1, B.2, and B.3) are repeated (with the same data until updated) in every frame, while each of the 25 frames in a complete message contains a different page of subframe 4 and 5. The first word of every subframe is a 30-bit encrypted telemetry (“TLM”) word (for users of the PPS) and the second word is a 30-bit Hand Over Word (“HOW”) generated by the SV. Therefore the HOW is transmitted every 6 seconds. The first seventeen bits of the HOW are the seventeen most significant bits of the TOW to aid the user’s receiver to rapidly lock on to the signal data (see Figure C.3). The actual TOW at the start of the next subframe is obtained by multiplying the TOW count contained in the HOW by 4.⁴ Hence the user knows the exact GPS time when the bit transition at the start of the next subframe occurs. At the end of the week, the TOW is reset to zero, and the frame pattern is repeated, starting with

⁴This comes from [1], but easier to see is each subframe is $6 = 100_2$ seconds in length, so 100_2 added to the truncated TOW (two least significant bits truncated) at the start of each subframe will yield the TOW at the start of the next subframe.

page one of subframe four and five, regardless of what page was last transmitted. Note that all SVs are transmitting in unison, and the TOW value in each SV's HOW is the same in each corresponding cycle. Thus we can compare arrival times of messages sent in the same cycle.

3.2 The User Segment

The GPS provides the means to solve two basic problems:

1. Given a user at a known position and the position of SV(s), determine Universal Coordinated Time (U.S. Naval Observatory), i.e., what time is it?
2. Given the position of four (or more) SVs, determine the user's location, i.e., where am I?

In the first problem, the distance from the user to each SV can be calculated to yield the direct propagation delay contributed by this distance. Adding the equipment group delay and the estimated ionospheric and tropospheric delays to the direct propagation delay yields the delay time to be added to (adjusted) GPS time, which then can be related to UTC (USNO).

In the second problem, the user can measure the clock bias needed to bring about a consistent position solution (x, y, z, t) to the system of equations incorporating an unknown delay factor added to the propagation time from the known positions of four (or more) SVs. (If the user is at a known elevation, e.g., at sea level, the signals from three SVs will suffice to solve for the ellipsoidal coordinates (λ, ϕ, t) .) This thesis will address this second problem.

Henceforth we shall use the symbols from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*. Refer to Appendix A for definitions of data as used in the Government GPS publications. In order to make use of the data in the NAV message, the values of the constants to be used and the number of significant figures are prescribed in *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*.

The sensitivity of position to the angular parameters is on the order of 10^8 meters/semicircle, and to the angular rate parameters is on the order of 10^{12} meters/semicircle/second. Because of this extreme sensitivity to angular perturbations, the value of π used in the curve fit is given here [1].

The WGS-84 parameters and the value of π prescribed are:

- the official WGS-84 speed of light, $c = 2.99792458e8$ meters/second

- the official WGS-84 value of the Earth’s universal gravitational parameter $\mu = 3.986005 \times 10^{14}$ meters³/sec²,
- the official WGS-84 value of the Earth’s rotation rate $\dot{\Omega}_e = 7.2921151467 \times 10^{-5}$ radians/sec., and
- $\pi = 3.1415926535898$.

3.2.1 Calculating Delays

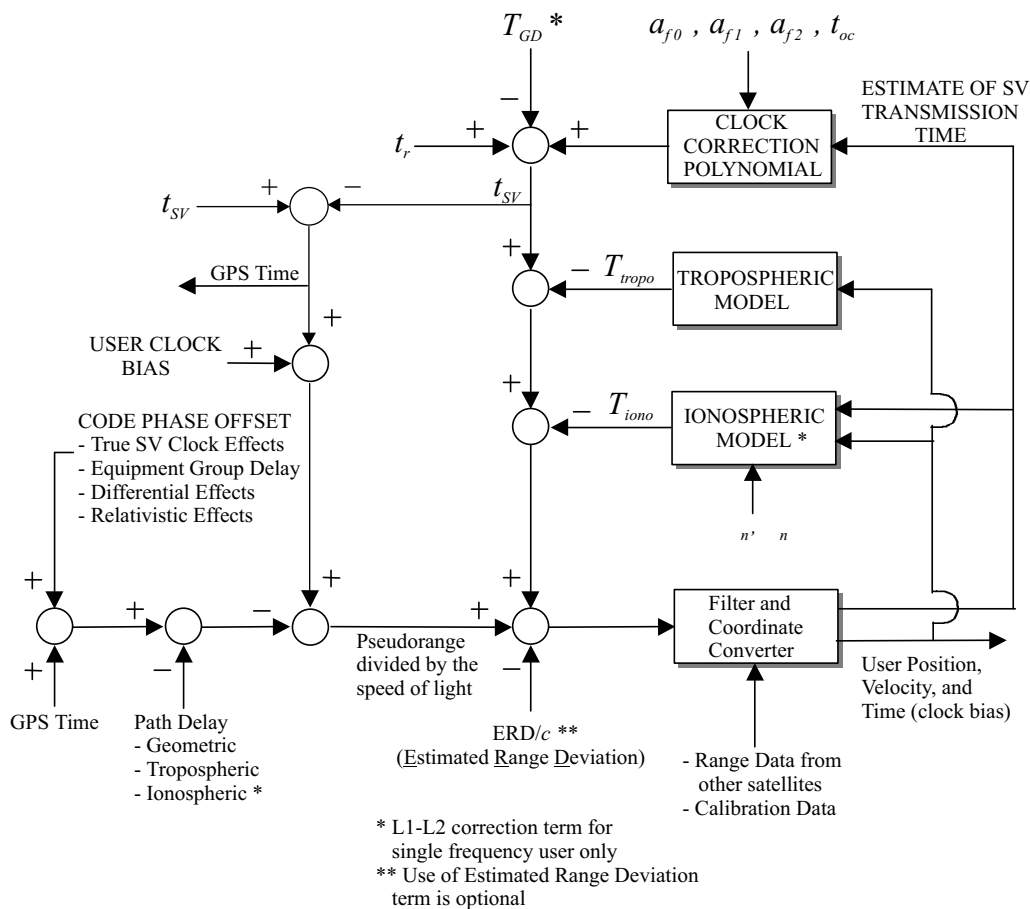


Figure 3.2: Application Of Correction Parameters. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*.)

A user’s clock must be synchronized with GPS time in order to calculate the time delay (or path delay) between the transmission of GPS time by the SV

and reception of that time by the user. The path delay multiplied by the speed of light yields the distance between the user and the SV which is then used in the calculation of the user's position. But the GPS time decoded from the NAV message received by a user is offset from the actual GPS time by several factors which must be accounted for.

- The SV clock drifts with respect to GPS time, and that offset is measured by the Control Segment. Correction for drift is accomplished by the user with a clock correction polynomial, the coefficients of which are determined by the Control Segment and provided to the user in the NAV message.
- Group delay, the time delay between the SV clock and the SV antenna is accounted for by subtracting the measured group delay, T_{GD} , in seconds, which is also provided in the NAV message by the Control Segment.
- Relativistic effects, caused by the relative motion between the user and the satellite and the Earth's gravitational field, are calculated by the user.
- Ionospheric delay, caused by refraction and delay of the radio signals as they pass through the ionosphere, is calculated by the user with a mathematical model provided.
- Tropospheric delay, caused by signal refraction in passage through the troposphere, may be accounted for at the user's discretion.
- Estimated Range Deviation ("ERD"), a calculated distance resulting from errors involved with the least squares fit of orbital parameters calculated by the Control Segment (which may be expressed in terms of time), is provided by the Control Segment in the NAV message. This delay equivalent may be accounted for optionally by the user.

Figure 3.2 shows the algorithm applied to each pseudorange to solve the position problem. We start with corrections to GPS time. GPS system time (at the time of transmission by the SV), t , is calculated in the first two summations,

$$t = t_{sv} - (\Delta t_{sv})_{L1},$$

where t_{sv} is the effective SV code phase time at message transmission time in seconds, and

$$(\Delta t_{sv})_{L1} = a_{f0} + a_{f1}(t - t_{oc}) + a_{f2}(t - t_{oc})^2 + \Delta t_r - T_{GD}.$$

Note that the previous two equations are coupled, i.e., t appears in both. The second equation is relatively insensitive to t , and so may be approximated by t_{sv}

[1]. The clock correction polynomial coefficients, a_{f0}, a_{f1}, a_{f2} , and t_{oc} , the clock data reference time, are calculated by and provided in subframe number one (see Figure B.1) by the Control Segment. Recall that T_{GD} is the group delay differential in seconds. The relativistic correction term, Δt_r , is defined

$$\Delta t_r = Fe\sqrt{a} \sin E_k,$$

where E_k is the eccentric anomaly, A is the length of the semi-major axis of the orbit, F is defined

$$F = \frac{-2(\mu)^{1/2}}{c^2},$$

and μ and c are defined above.

The summation node on the lower left corner of Figure 3.2 depicts the SV, and proceeding to the right, GPS time suffers the path delay transiting the ionosphere and troposphere. Then in the next summation pseudorange is calculated. The input to the pseudorange calculation from above is the user's time, offset from GPS time by the amount of receiver clock bias. The pseudorange is then divided by the speed of light, yielding the pseudorange in seconds which is applied to the next summation for correction due to delays.

3.2.2 Ionospheric Delay Model

A specific tropospheric model is not prescribed in *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*, but an ionospheric model is provided and reproduced here ⁵ without elaboration. The ionospheric delay estimation, T_{iono} is given by

$$T_{iono} = \begin{cases} F \left[5.0 \times 10^{-9} + (AMP) \left(1 - \frac{x^2}{2} + \frac{x^4}{24} \right) \right] & , |x| < 1.57 \\ F(5.0 \times 10^{-9}) & , |x| \geq 1.57 \end{cases} \text{ seconds,}$$

where

$$AMP = \begin{cases} \sum_{n=0}^3 \alpha_n \phi_m^n, & AMP \geq 0 \\ \text{if } AMP < 0, & AMP = 0 \end{cases} \text{ seconds,}$$

$$x = \frac{2\pi(t - 50400)}{PER} \text{ radians,}$$

$$PER = \begin{cases} \sum_{n=0}^3 \beta_n \phi_m^n, & PER \geq 72,000 \\ \text{if } PER < 72,000, & PER = 72,000 \end{cases} \text{ seconds,}$$

$$F = 1.0 + 16.0(0.53 - E)^3,$$

⁵This model is taken directly, with some paraphrasing, from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*.

and α_n and β_n , $n = 0, 1, 2, 3$ are transmitted in subframe 4, page 18. Additionally,

$$\begin{aligned}\phi_m &= \phi_i + 0.064 \cos(\lambda_i - 1.617) \\ \lambda_i &= \lambda_u + \frac{\psi \sin A}{\cos \phi_i} \\ \phi_i &= \begin{cases} \phi_u + \psi \cos A \text{ (semi-circles),} & |\phi_i| \leq 0.416 \\ \text{if } \phi_i > 0.416, & \phi_i = +0.416 \\ \text{if } \phi_i < -0.416, & \phi_i = -0.416 \end{cases} \\ \psi &= \frac{0.00137}{E + 0.11} - 0.022 \\ t &= (4.32 \times 10^4)\lambda_i + \text{GPS time in seconds,}\end{aligned}$$

where $0 \leq t < 86400$. Therefore, if $t \geq 86400$ seconds, subtract 86400 seconds, and if $t < 0$ seconds, add 86400 seconds. The terms used in the model are defined as follows:

- SV transmitted terms
 - α_n are the terms of a cubic equation representing the amplitude of the vertical delay (4 coefficients, 8 bits each),
 - β_n are the coefficients of a cubic equation representing the period of the model (4 coefficients, 8 bits each).
- Receiver Generated terms
 - E is the elevation angle between the user and the SV in semi-circles,
 - A is the azimuth angle between the user and satellite, measured clockwise positive from true North in semi-circles,
 - ϕ_u is the user's geodetic latitude in semi-circles, WGS-84,
 - λ_u is the user's geodetic longitude in semi-circles, WGS-84, and
 - GPS time is the receiver computed GPS system time.
- Computed terms are
 - x is the phase in radians,
 - F is the obliquity factor,
 - t is local time in seconds,
 - ϕ_m is the geomagnetic latitude of the Earth projection of the ionospheric intersection point in semi-circles (mean ionospheric height assumed 350 kilometers),

- λ_i is the geodetic longitude of the Earth projection of the ionospheric intersection point in semi-circles,
- ϕ_i is the geodetic latitude of the Earth projection of the ionospheric intersection point in semi-circles ⁶, and
- ψ is the Earth’s central angle between user position and Earth projection of ionospheric intersection point in semi-circles.

Note that the term E should not be confused with the Eccentric Anomaly, E_k , and the term A used here is *not* the semi-major axis of the orbit, as used elsewhere in the same publications. The ionospheric prediction model stands on its own. Note also that the calculation depends on knowing the user’s position (latitude and longitude), hence the position problem must be solved first without an ionospheric delay estimate, then the delay calculation may be performed, and by iteration the solution may be refined. As stated above, the ionospheric delay estimate is the largest source of uncertainty in the position solution.

3.2.3 Calculating SV Position

The box in the lower right corner of Figure 3.2, labelled *Filter and Coordinate Converter*, is expanded into the algorithm prescribed in *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200* (and also in *Global Positioning System Standard Positioning Service Signal Specification*) for the solution of SV position and is shown in the Appendix, Figure C.1 and Figure C.2. Use the definitions of ephemeris parameters in Figure A.1 to correlate symbols used in section 2.2. Note that

... the orbital parameters [are] typical of Keplerian orbital parameters; it shall be noted, however, that the transmitted parameter values are such that they provide the best trajectory fit in Earth-Centered, Earth-Fixed (ECEF) coordinates for each specific fit interval. The user shall not interpret intermediate coordinate values as pertaining to any conventional coordinate system [1].

The descriptions of the Keplerian orbital parameters derived in the preceding sections of this thesis are based on orbits taken as a whole, the orbits calculated in the following algorithm are based on a least-squares fit of measured

⁶This definition and the preceding one differ between *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200* (shown here) and *Global Positioning System Standard Positioning Service Signal Specification*. I assume the former publication is correct, and the latter is in error on this point.

data to provide the accuracy required in the GPS over a short interval (typically four hours).

The algorithm (Figure C.1) begins by retrieving \sqrt{A} from subframe 2, words 8 and 9 and squaring this value to obtain A , the semi-major axis. With A and μ the mean motion is computed from the next formula, equivalent to (2.29).

Next the elapsed time since the reference ephemeris time is calculated. This is the term $(t - \tau)$ of (2.30). Since SV time is a binary count of X1 epochs, one must account for the weekly resetting of the TOW count as described in the note at the bottom of Figure C.1.

The value of Δn , the mean motion difference from computed value, is retrieved from subframe 2, word 4, and we add this to the computed mean motion to obtain the corrected mean motion.

Mean anomaly is computed next, after retrieving the mean anomaly at reference time, M_0 , from subframe 2, words 4 and 5, and adding the mean anomaly *since* reference time, (2.30), the product of the elapsed time from ephemeris reference epoch and the corrected mean motion.

Having the mean anomaly in hand, we use Kepler's equation (2.38) to calculate the eccentric anomaly by iteration.

Now given the eccentric anomaly, we can calculate the true anomaly. Note that the equation for true anomaly shown in Figure C.1, after reduction and some manipulation, is equivalent to (2.32).

After extracting the argument of perigee, ω , from subframe 3, words 7 and 8, we construct the argument of latitude by adding the argument of perigee to the true anomaly. Note that the argument of latitude is the angle the radius vector of the SV has swept since the right ascension of the ascending node, and thus is proportional to the latitude of the SV as indicated in section 2.2.1. When the argument of latitude is zero, the latitude is zero, and when the argument of latitude is $\pi/2$ radians, the latitude is the orbital inclination, i.e., nominally 54° .

We calculate the perturbations of the orbital parameters next (see Figure C.2). The amplitudes of the sine and cosine harmonic correction terms are obtained from subframes 1 and 2 (see Figures B.1 and B.2). Note that the form of the corrections to the argument of latitude, the orbit radius, and the inclination are of the same form as (2.48).

The corrections are next added to the argument of latitude, the orbit radius, and the inclination to obtain the corrected orbital parameters.

The corrected orbital parameters are then used to find the SV position in the orbital plane. Note in particular that the radius is calculated from (2.31) using the eccentric anomaly, and there is an assumption implicit in

$$x'_k = r_k \cos u_k \text{ and } y'_k = r_k \sin u_k$$

that the center of the orbit, the Earth's center of mass, is also the focus about which the SV orbits. This is not an unreasonable assumption, given the nearly circular orbits of the SVs, and the corrections (particularly the radius correction) fitted to the orbit over short periods of time.

In the corrections to the longitude of the ascending node, the first term, Ω_0 , is the longitude of the ascending node of the orbit plane at weekly epoch. To this we add the change in the right ascension since the time of ephemeris epoch ($\dot{\Omega}t_k$), subtract the earth's rotation angle since the time of ephemeris epoch ($\dot{\Omega}_e t_k$), and then subtract the earth's rotation angle since the time of ephemeris.

Finally, the polar positions in the orbital plane (which are referenced to the x axis in the x, y plane) are rotated about the z axis through the angle of the corrected longitude of the ascending node, Ω_k , and then rotated through the angle of inclination, i_k . At this point we have obtained the Cartesian coordinates (x_s, y_s, z_s) of the SV in the ECEF reference frame.

3.2.4 Calculating User Position

Now that we have calculated the GPS time offset between the SV transmitted time and actual GPS time, and we have calculated the position of the SV, we are in position to calculate the path delay. After obtaining the path delay, the user's clock will be synchronized with GPS time and the distance from the user to the SV will be known, enabling the calculation of the user's position.

At this point, it is beneficial to take a high-level look at the problem of determining the actual time delay of the delayed GPS time signal. If the SV is directly overhead, referenced to a user on the surface of the Earth, the SV will be about

$$26560 - 6378 = 20182$$

kilometers away. Electromagnetic waves propagate at about 300×10^6 meters/second, so the signal from the SV directly overhead will be delayed approximately 67 milliseconds. Electromagnetic waves at this frequency propagate pretty much in straight lines, with some refraction introduced by the ionosphere and troposphere. Therefore, a user will receive signals only from SVs above the horizon (in fact GPS receiver antennas typically discriminate against signals close to the horizon). Then the distance from a user to the SV on the horizon is

$$\frac{6378}{\tan\left(\sin^{-1}\frac{6378}{26560}\right)}$$

or about 25783 kilometers, representing somewhat less than 87 milliseconds signal transit time. So when a GPS receiver decodes the time from the SV, we know that it is (roughly) no less than 67 milliseconds and no more than 87 milliseconds old. Then upon receiving a time signal from one SV, we know GPS

time within about 20 milliseconds. Receiving time signals from four SVs (the number we need to solve for four unknowns) or more will narrow this window because we also know that the first signal to arrive has been delayed no less than 67 milliseconds, and the last signal to arrive has been delayed no more than 87 milliseconds. For example, if the difference in arrival times between the first message and the last message is 16 milliseconds, we know the delay of all the messages is somewhere between 67 and 71 milliseconds and the receiver's clock must be biased in this neighborhood to produce appropriate pseudoranges. What is more, we can measure the *difference* in arrival times of the signals, say for example, to the nearest 10 nanoseconds (roughly 3 meters in distance). The receiver clock bias could then be varied until a consistent position solution is obtained from the pseudoranges within 3 meters (10 nanoseconds), yielding the apparent position of the receiver in the neighborhood of 3 meters. Obviously the accuracy of this process would depend on the ability to resolve time differences in signal arrival times and resolve the receiver's clock.

We know then that the user's implied position, (x, y, z) , lies on a sphere of radius equal to the pseudorange p centered at the SV's position, (x_s, y_s, z_s) , $s = 1, 2, 3, 4$. Therefore

$$\begin{aligned} p_s &= \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2} \\ p_s^2 &= (x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2 \\ &= x^2 + y^2 + z^2 + x_s^2 + y_s^2 + z_s^2 - 2xx_s - 2yy_s - 2zz_s, \end{aligned}$$

so that

$$p_s = (x^2 + y^2 + z^2) + r_s^2 - 2xx_s - 2yy_s - 2zz_s, \quad (3.1)$$

where

$$r_s = \sqrt{x_s^2 + y_s^2 + z_s^2}$$

is the distance from the origin to SV number s . Note that the square of the pseudorange in (3.1) varies directly with the square of the distance from the origin to the user, $(x^2 + y^2 + z^2)$, considered together as another variable. Logic dictates an inverse relationship between the pseudoranges and the distance from the origin to the user; therefore the system of equations is consistent only when the computed value of (x, y, z) (considering $(x^2 + y^2 + z^2)$ as a fourth variable) yields a distance squared from the origin to the user equal to the fourth variable. Therefore, let $(c\tau)^2$, where c is the speed of light and τ is the receiver time offset parameter, be equivalent to the distance from the origin to the user. We decouple $(c\tau)^2$ from the variables (x, y, z) so that (3.1) becomes

$$p^2 = (c\tau)^2 + r_s^2 - 2xx_s - 2yy_s - 2zz_s. \quad (3.2)$$

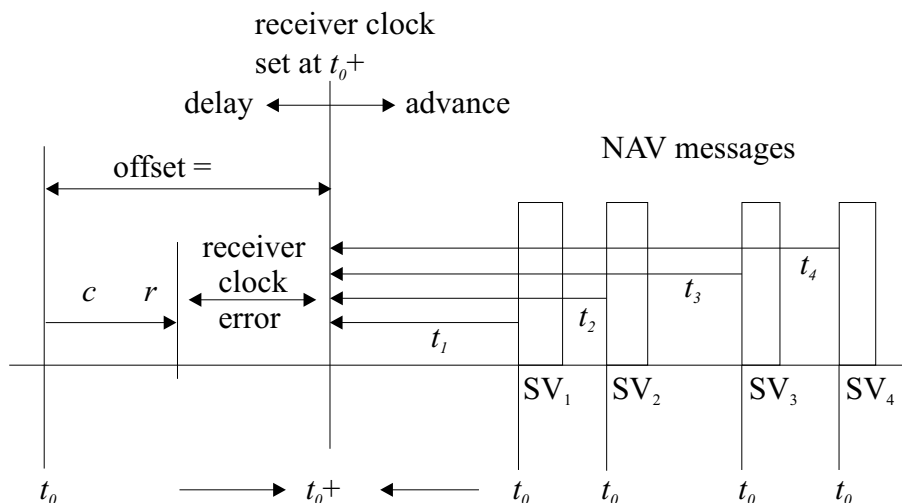


Figure 3.3: Receiver Clock Offset.

The unknowns in (3.2) are now x, y, z and $(c\tau)^2$; and p is varied by adjusting the value of the receiver clock offset so that

$$\sqrt{x^2 + y^2 + z^2} = (c\tau).$$

Figure 3.3 is a schematic depiction of the timing relationships among the signal delays and the receiver clock offset. Each SV initiates a NAV message (depicted on the right side of diagram) at GPS time t_0 ; and each signal is delayed t_s seconds before reaching the receiver. At t_0 another hypothetical radio signal launched from a distance of $c\alpha\tau$ meters away (depicted on the left side of diagram) would arrive at the receiver at the same time as the NAV message from the nearest SV. The constant α is undetermined. Since τ is to be determined, we interpret τ as the receiver clock offset when the GPS time has been accurately determined by the receiver. Thus we can vary the pseudoranges (by varying the receiver clock offset) until we have a consistent solution for x, y, z, τ .

We can find the sensitivity of position determination to timing errors by taking the partial derivative of p_s with respect to $c\tau$ in (3.2),

$$\frac{\partial p_s}{\partial(c\tau)} = \frac{(c\tau)}{p_s},$$

and because $p_s = ct_s$, we know that

$$\frac{\partial p_s}{\partial(c\tau)} = \frac{\tau}{t_s}.$$

After finding the partial derivatives of p with respect to $x, y, z, c\tau$ in (3.2) we form the Jacobian matrix

$$\begin{bmatrix} \frac{x-x_1}{p_1} & \frac{y-y_1}{p_1} & \frac{z-z_1}{p_1} & \frac{\tau}{t_1} \\ \frac{x-x_2}{p_2} & \frac{y-y_2}{p_2} & \frac{z-z_2}{p_2} & \frac{\tau}{t_2} \\ \frac{x-x_3}{p_3} & \frac{y-y_3}{p_3} & \frac{z-z_3}{p_3} & \frac{\tau}{t_3} \\ \frac{x-x_4}{p_4} & \frac{y-y_4}{p_4} & \frac{z-z_4}{p_4} & \frac{\tau}{t_4} \end{bmatrix}. \quad (3.3)$$

The Jacobian matrix contains all the information about the sensitivities of the position solution to variations in the pseudoranges and time offset. It is not dependent upon any coordinate system, and hence the sensitivities of the position solution are embodied in the four invariants of this matrix.

Proceeding with the solution of position, let

$$\mathbf{A} = \begin{bmatrix} 2x_1 & 2y_1 & 2z_1 & -1 \\ 2x_2 & 2y_2 & 2z_2 & -1 \\ 2x_3 & 2y_3 & 2z_3 & -1 \\ 2x_4 & 2y_4 & 2z_4 & -1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \\ z \\ (c\tau)^2 \end{bmatrix}, \mathbf{r}_s = \begin{bmatrix} r_1^2 \\ r_2^2 \\ r_3^2 \\ r_4^2 \end{bmatrix}, \text{ and } \mathbf{p}_s = \begin{bmatrix} p_1^2 \\ p_2^2 \\ p_3^2 \\ p_4^2 \end{bmatrix}.$$

Then

$$\mathbf{A}\mathbf{x} = \mathbf{r}_s - \mathbf{p}_s,$$

and therefore

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{r}_s - \mathbf{A}^{-1}\mathbf{p}_s, \quad (3.4)$$

assuming that \mathbf{A} is non-singular. (\mathbf{A} can become singular with particular user-SV geometries, but that condition should exist only for one observation at a time, given the rapidly changing nature of the components of the matrix.)

A possible method ⁷ of solving for the user's position (and determining the receiver clock offset) following the logic set out above, is as follows. Let the signal time delays, $t_n, n = 1, 2, 3, 4$ from the SVs be arranged from closest to farthest away, where the time delay of the signal from the nearest SV is arbitrarily assumed to be t_1 (and the time delays from the other SVs maintain their relationship with t_1). Then

$$\mathbf{p}_s = \begin{bmatrix} t_1^2 \\ t_2^2 \\ t_3^2 \\ t_4^2 \end{bmatrix} c^2.$$

Consider a convergent sequence of delay times (pseudoranges/ c) that begins at the arbitrarily chosen t_1 and converges to the value of the actual time delay of

⁷This is not the only way of approaching a solution, but it relates to the intuitive approach outlined herein.

the signal from the nearest SV to the user. We let the sequence be $\{s_n\}, n = 1, 2, \dots$, and successively set $t_1 = s_n$ for $n = 1, 2, 3, \dots$. Then \mathbf{p}_s will converge to the actual ranges (squared) to the SVs.

Let the first term of the sequence, s_1 , be less than the minimum delay anticipated from the nearest SV, i.e., 67 milliseconds. Assume the signal from the farthest SV observed is delayed less than δ milliseconds from that of the nearest SV, where δ is the delay rounded up to the next integer number of milliseconds. Thus the range of possible delays is 67 to $87 - \delta$ milliseconds. Let the difference, $20 - \delta$, be designated d . Successive terms of the convergent sequence will be such that $|s_{n+1} - s_n| = d/2^n$. Thus the second term of the sequence, s_2 , be half the difference between the minimum predicted delay and the maximum predicted delay, i.e., $67 + d/2$. The second member of the sequence is therefore within $d/2$ milliseconds of the actual delay. Then $s_2 = t_1$ is used in (3.4). The third term of the sequence will be either $67 + d/2 + d/4$ milliseconds or $67 + d/2 - d/4 = 67 + d/4$ milliseconds and will be within $d/2^2$ of the actual delay. The value of the third term depends on the sign of the steering function $\sqrt{x^2 + y^2 + z^2} - (c\tau)$; a change of sign reverses the direction (increase/decrease time) of the sequence. Then $s_3 = t_1$ is used in (3.4), and the sequence continues for n iterations until $d/2^n$ is less than the acceptable error. For example, if $d = 4$ milliseconds and $n = 20$, the delay has been estimated within about 4 nanoseconds, and the range has been estimated within about 1.2 meters, neglecting ionospheric and tropospheric delays. The final values of the ranges are then derived from the last iteration of (3.4) as are the user's Cartesian coordinates in the ECEF reference frame and the receiver clock offset τ . At this point, the ionospheric and tropospheric delays can be calculated and subtracted from the delays to yield the actual ranges and times. Also, once the user's position is calculated, the last calculated position can be used as the estimate of the user's next position. The next calculation, using updated SV positions, can thus begin with a very close estimate of pseudorange and yield the desired result with fewer iterations.

Finally, with the user's Cartesian coordinates in the ECEF reference frame one determines h from (1.6), longitude from (1.4) and the latitude ϕ from (1.3) as outlined in section 1.1.

3.2.5 Dilution of Precision

While not directly a part of determining position using the GPS, a little reflection on the nature of the problem of solving for position by trilateration using satellites leads one to realize that one's confidence in the solution relies on the geometry of the problem. For example, given that the orbits of the GPS satellites are inclined about 54° with respect to the equatorial plane, a GPS user

located at latitude 54°N would observe only SVs from directly overhead and to the south. Farther north, an observer would see only satellites low on the horizon, including those to the north when the user is far enough north. One can imagine other scenarios, e.g., all SVs close to the horizon, all SVs high overhead, each geometry creating different uncertainties. One geometry might lead to more uncertainty in elevation, e.g., satellites low in the sky, while another might lead to uncertainty in horizontal position, e.g., all satellites high in the sky. Some reflection leads one to conjecture that the uncertainty is related to the volume of the body enclosed by the vectors from observation site to the satellites, and one should attempt to maximize this volume. This is indeed the case. Specifically, the volume we would like to maximize is that of the solid described by the site-satellite vectors and the points of intersection with the unit sphere centered at the user position [8].

The measure of satellite geometry that relates to the quality of the SV geometry is *dilution of precision*, *DOP*. The starting point is the Jacobian matrix (3.3). There are four invariants associated with this 4 by 4 matrix, the most commonly known is the determinant. The invariants of this Jacobian matrix are interpreted as the *HDOP*, the horizontal dilution of precision; the *VDOP*, the vertical dilution of precision; the *GDOP*, the geometric dilution of precision; and the *PDOP*, position dilution of precision. Additionally, a *TDOP*, time dilution of precision, is derived. It should be noted that most receivers now calculate the DOP numbers for the satellites in view, and use the four yielding the best DOP numbers.

One does not need to be able to calculate the DOP numbers to determine position from the GPS, but their calculation is a direct offshoot of the position solution, i.e., (3.3). Anyone who wishes to know more about the GPS will find DOP to be one of the next subjects to pursue, as the accuracy and the confidence one may have in the position calculations hinge on DOP.

4 Final Thoughts

Calculating position by trilateration using the GPS covers a broad range of subjects in mathematics and physics, and of necessity some parts could only be touched on lightly in this primer. They are subjects in their own right and worthy of investigation for their own sake.

Several areas that bear on the GPS were not addressed in this thesis, so I mention a few here. Because ionospheric effects are the greatest source of error in the use of the GPS, several methods are utilized to overcome the effects of ionospheric refraction. The PPS makes use of the fact that refraction is a function of frequency, and the use of two radio frequency carriers provides the means to eliminate the ionospheric effects. Differential GPS makes use of the fact that the delays are the same for receivers located in the same vicinity, and therefore the *difference* between simultaneous measurements does not suffer from the ionospheric effects. Millimeter accuracy is achievable with differential GPS. The Federal Aviation Administration's Wide Area Augmentation System ("WAAS") measures the ionospheric effects and relays correction factors to a geostationary satellite, which are in turn transmitted to GPS users over a wide area. Several receivers are available that make use of the WAAS.

Accuracy of the calculated position is also not addressed in this thesis. The Government defines three measures of accuracy, defined on the basis of the SV constellation [2]. Because ionospheric effects, the choice of receiver (the manufacturer's methods of solution), and the varying DOP are beyond the control of the Government, one must account for these after determining the effects described in the *Global Positioning System Standard Positioning Service Performance Standard* [2].

Lastly, I must say that I have been fascinated by the possibilities of using satellites for any number of things since the launch of the first Sputnik. GPS, in particular, attracted my interest since it is a technology designed in part for the public, one that you can hold in your hand and see the direct result of the application of satellite technology. I knew that one day I would learn, from a mathematical perspective, how it works from start to finish. This thesis is part of the result of that (unfinished) challenge.

A Ephemeris Definitions

M_0	Mean Anomaly at Reference Time
n	Mean Motion Difference From Computed Value
e	Eccentricity
$(A)^{1/2}$	Square Root of the Semi-Major Axis
$(\text{OMEGA})_0$	Longitude of Ascending Node of Orbit Plane at Weekly Epoch
i_0	Inclination Angle at Reference Time
	Argument of Perigee
OMEGADOT	Rate of Right Ascension
IDOT	Rate of Inclination Angle
C_{uc}	Amplitude of the Cosine Harmonic Correction Term to the Argument of Latitude
C_{us}	Amplitude of the Sine Harmonic Correction Term to the Argument of Latitude
C_{rc}	Amplitude of the Cosine Harmonic Correction Term to the Orbit Radius
C_{rs}	Amplitude of the Sine Harmonic Correction Term to the Orbit Radius
C_{ic}	Amplitude of the Cosine Harmonic Correction Term to the Angle of Inclination
C_{is}	Amplitude of the Sine Harmonic Correction Term to the Angle of Inclination
t_{oe}	Reference Time Ephemeris
IODE	Issue of Data (Ephemeris)

Figure A.1: Ephemeris Definitions. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200.*)

B Subframes 1, 2 and 3

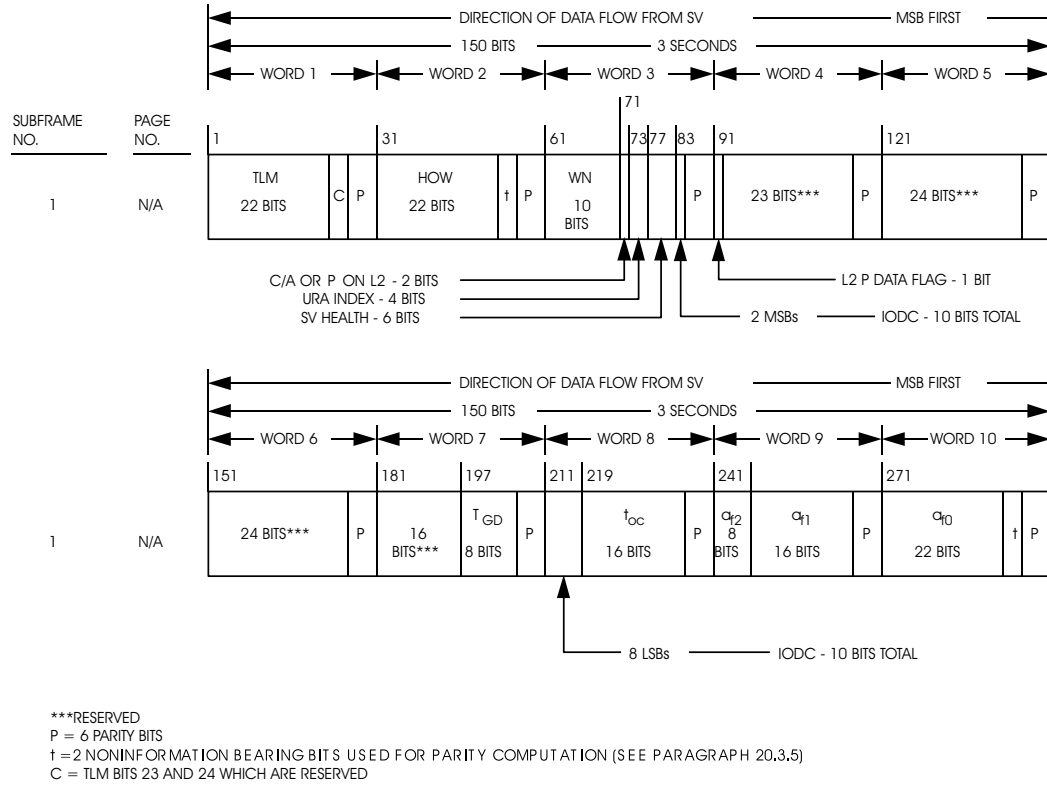


Figure B.1: Contents Of Subframe 1. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200.*)

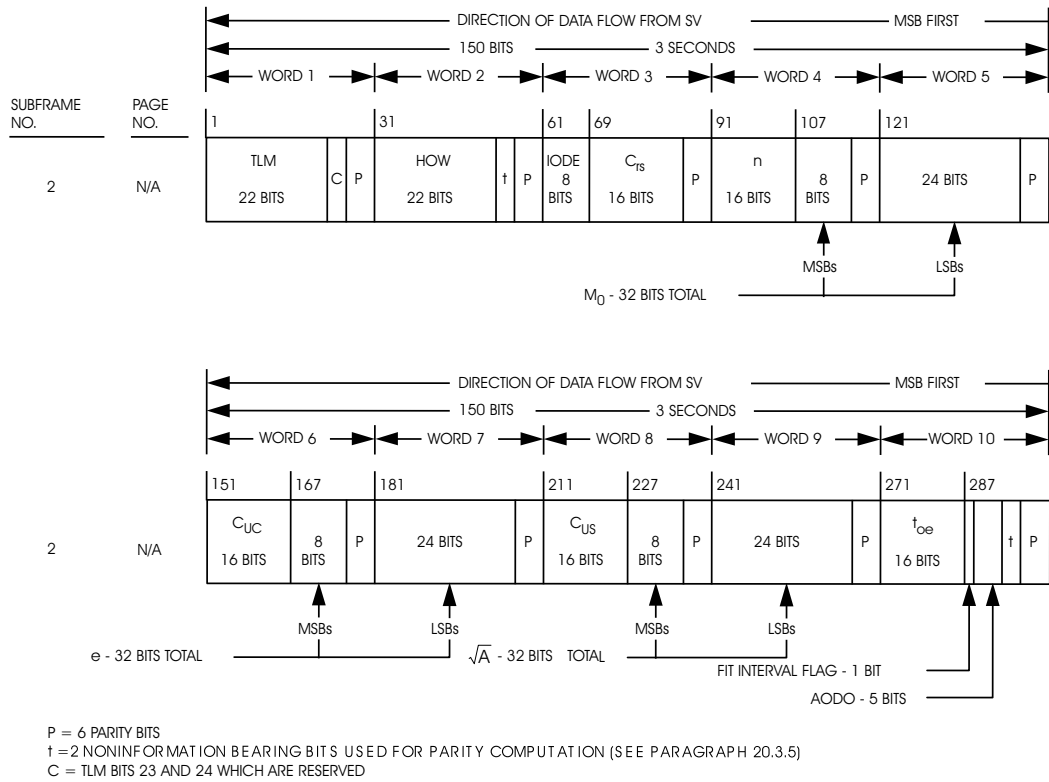
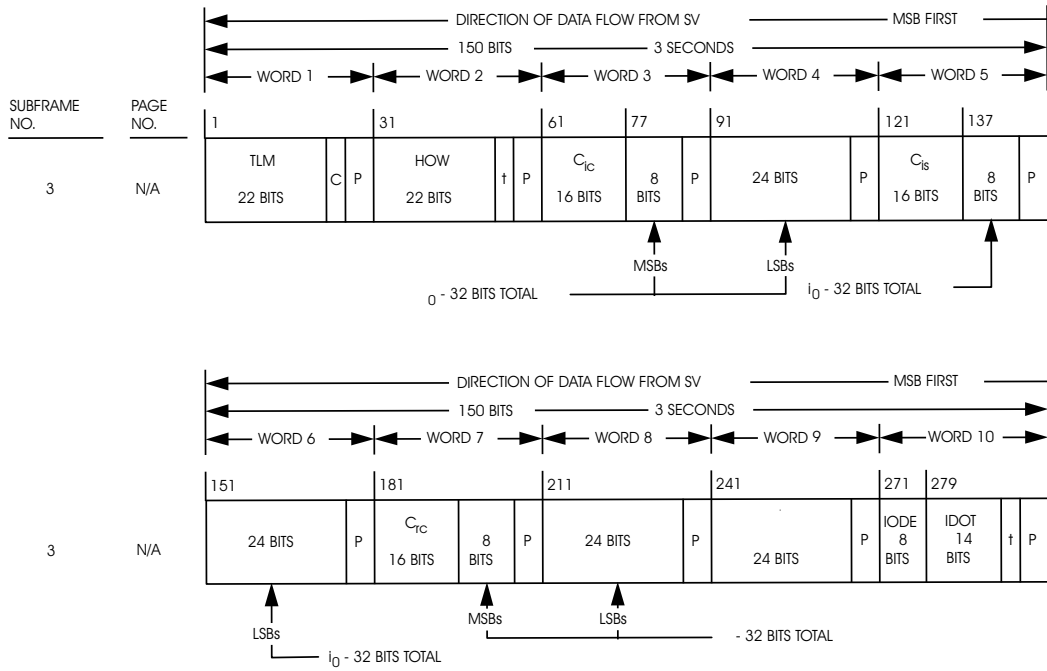


Figure B.2: Contents Of Subframe 2. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200.*)



P = 6 PARITY BITS
 † = 2 NONINFORMATION BEARING BITS USED FOR PARITY COMPUTATION (SEE PARAGRAPH 20.3.5)
 C = TLM BITS 23 AND 24 WHICH ARE RESERVED

Figure B.3: Contents Of Subframe 3. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200.*)

C ECEF Algorithm And HOW

$= 3.986005 \times 10^{14} \text{meters}^3/\text{sec}^2$	WGS 84 value of the earth's universal gravitational parameter for GPS user
$\omega_e = 7.2921151467 \times 10^{-5} \text{rad/sec}$	WGS 84 value of the earth's rotation rate
$A = (\sqrt{A})^2$	Semi-major axis
$n_0 = \sqrt{\frac{\mu}{A^3}}$	Computed mean motion (rad/sec)
$t_k = t - t_{oe}^*$	Time from ephemeris reference epoch
$n = n_0 + \dot{n}$	Corrected mean motion
$M_k = M_0 + nt_k$	Mean anomaly
$M_k = E_k - e \sin E_k$	Kepler's Equation for Eccentric Anomaly (may be solved by iteration)(radians)
$\nu_k = \tan^{-1} \frac{\sin E_k}{\cos E_k}$	True Anomaly
$= \tan^{-1} \left[\frac{\sqrt{1-e^2} \sin E_k / (1-e \cos E_k)}{(\cos E_k - e) / (1-e \cos E_k)} \right]$	
$E_k = \cos^{-1} \left[\frac{e + \cos \nu_k}{1 + e \cos \nu_k} \right]$	Eccentric Anomaly
$\lambda_k = \nu_k + \omega_k$	Argument of Latitude

t is GPS system time at time of transmission, i.e., GPS time corrected for transit time (range/speed of light). Furthermore, t_k shall be the actual total time difference between the time t and the epoch time t_{oe} and must account for beginning or end of week crossovers. That is, if t_k is greater than 302,400 seconds, subtract 604,800 seconds from t_k . If t_k is less than -302,400 seconds, add 604,800 seconds to t_k .

Figure C.1: ECEF Algorithm, Part A. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*.)

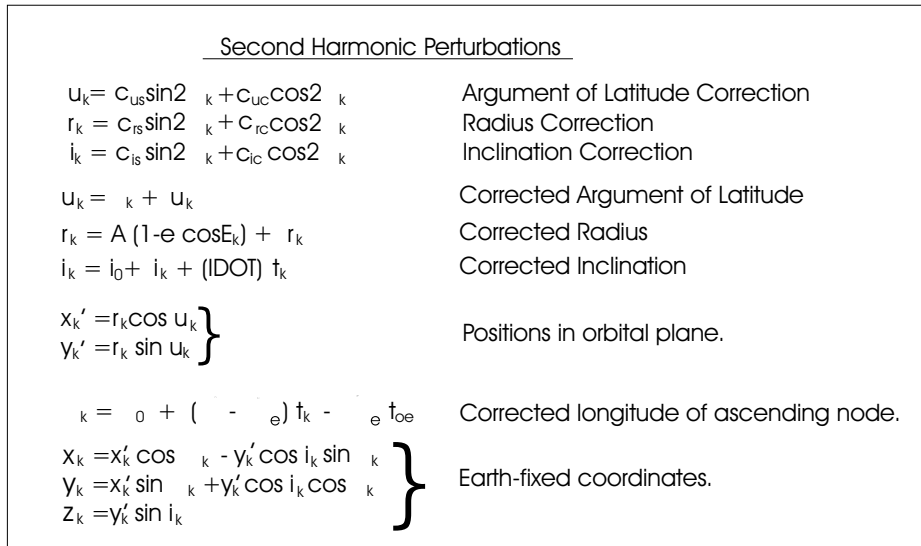


Figure C.2: ECEF Algorithm, Part B. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200.*)

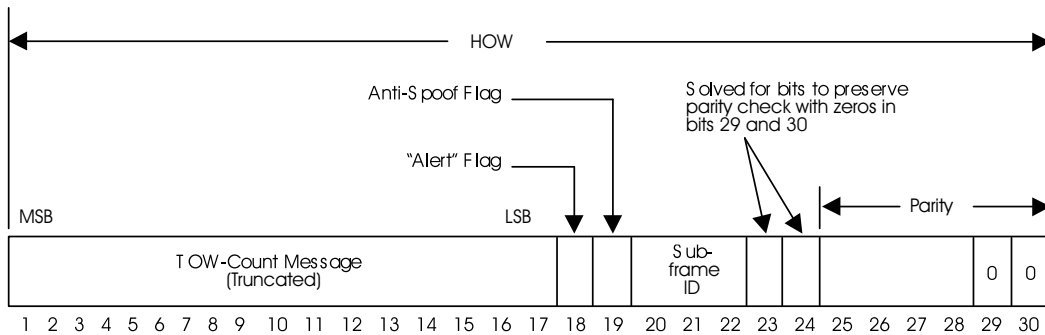


Figure C.3: HOW. (Adapted from *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200.*)

References

- [1] ARINC Research Corporation. *Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200*. U.S. Government, 2000.
- [2] Assistant Secretary of Defense. *Global Positioning System Standard Positioning Service Performance Standard*. U.S. Department of Defense, 2001.
- [3] W.H. Beyer. *CRC Standard Mathematical Tables*, 28th Edition. Boca Raton, Florida: CRC Press, Inc., 1987.
- [4] GPS Navstar Global Positioning System. *Global Positioning System Standard Positioning Service Signal Specification*. U.S. Government, 1995.
- [5] H.F. Davis. *Introduction to Vector Analysis*, Second Edition. Boston: Allyn and Bacon, Inc., 1970.
- [6] R.M. Green. *Spherical Astronomy*. Cambridge: Cambridge University Press, 1985.
- [7] D. Halliday and R. Resnick. *Fundamentals of Physics*, Second Edition. New York: John Wiley and Sons, 1981.
- [8] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins. *GPS Theory and Practice*, Third Revised Edition. New York: Springer-Verlag Wein, 1994.
- [9] L. Hogben. *Mathematics for the Million*. New York, London: W.W. Norton & Company, 1968.
- [10] V. Katz. *A History of Mathematics*. New York: Harper Collins College Publishers, 1993.
- [11] W.M. Kaula. *Theory of Satellite Geodesy*. Mineola, New York: Dover Publications, Inc., 2000.
- [12] A. Leick. *GPS Satellite Surveying*. New York: John Wiley & Sons Inc., 1990.
- [13] F.R. Moulton. *An Introduction to Celestial Mechanics*, Republication of the Second Revised Edition. New York: Dover Publications, Inc., 1914.

- [14] A.E. Roy. *Orbital Motion*, Third Edition. Bristol and Philadelphia: Adam Hilger, 1988.
- [15] C.L.Siegel and J.K. Moser. *Lectures on Celestial Mechanics*, Reprint of the 1971 Edition. Berlin, Heidelberg, New York: Springer-Verlag, 1995.
- [16] G. Strang and K. Borre. *Linear Algebra, Geodesy, and GPS*. Wellesley, Massachusetts: Wellesley-Cambridge Press, 1997.
- [17] V.G. Szebehely. *Adventures in Celestial Mechanics*. Austin, Texas: University of Texas Press, 1989.
- [18] L.G. Taff. *Celestial Mechanics*. New York: John Wiley and Sons, 1985.