COORDINATE SYSTEMS IN GEODESY

E. J. KRAKIWSKY
D. E. WELLS

May 1971

GEODESY AND GEOMATICS ENGINEERING
UNB

LECTURE NOTES
NO. 16

COORDINATE SYSTEMS IN GEODESY

E.J. Krakiwsky
D.E. Wells

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

May 1971
Latest Reprinting January 1998

PREFACE

In order to make our extensive series of lecture notes more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

TABLE OF CONTENTS

page

LIST OF ILLUSTRATIONS iv

LIST OF TABLES vi

1. INTRODUCTION 1

1.1 Poles, Planes and Axes 4
1.2 Universal and Sidereal Time 6
1.3 Coordinate Systems in Geodesy 7

2. TERRESTRIAL COORDINATE SYSTEMS 9

2.1 Terrestrial Geocentric Systems 9

2.1.1 Polar Motion and Irregular Rotation of the Earth 10
2.1.2 Average and Instantaneous Terrestrial Systems . 12
2.1.3 Geodetic Systems 17

2.2 Relationship between Cartesian and Curvilinear Coordinates 19

2.2.1 Cartesian and Curvilinear Coordinates of a Point on the Reference Ellipsoid 19
2.2.2 The Position Vector in Terms of the Geodetic Latitude 22
2.2.3 The Position Vector in Terms of the Geocentric and Reduced Latitudes 27
2.2.4 Relationships between Geodetic, Geocentric and Reduced Latitudes 28
2.2.5 The Position Vector of a Point Above the Reference Ellipsoid 28
2.2.6 Transformation from Average Terrestrial Cartesian to Geodetic Coordinates 31

2.3 Geodetic Datums 33

2.3.1 Datum Position Parameters 36
2.3.2 Establishment of a Datum 40
2.3.3 The North American Datum 42
2.3.4 Datum Transformations 45

2.4 Terrestrial Topocentric Systems 48

2.4.1 Local Astronomic System 50
2.4.2 Local Geodetic System 54

2.5 Summary of Terrestrial Systems 57

		page
3.	CELESTIAL COORDINATE SYSTEMS	61
3.1	The Ecliptic System	63
3.2	The Right Ascension System	65
3.3	The Hour Angle System	67
3.4	The Horizon System	71
3.5	Variations of the Right Ascension System	73
3.5.1	Precession and Nutation	74
3.5.2	Mean Celestial Systems	76
3.5.3	The True Celestial System	81
3.5.4	The Apparent Place System	83
3.5.5	The Observed Place System	84
3.6	Transformation between Apparent Celestial and Average Terrestrial Coordinate Systems	85
3.7	Summary of Celestial Systems	87
4.	THE ORBITAL COORDINATE SYSTEM	92
4.1	The Orbital Ellipse and Orbital Anomalies	92
4.2	The Orbital Coordinate System	96
4.3	Transformation from Orbital to Average Terrestrial System	98
4.4	Variations in the Orbital Elements	99
4.5	The Satellite Subpoint	99
4.6	Topocentric Coordinates of Satellite	101
5.	SUMMARY OF COORDINATE SYSTEMS	104
5.1	Terrestrial Systems	104
5.2	Celestial Systems	106
5.3	Duality Paradox in the Apparent and Observed Celestial Systems	107
5.4	The Connections between Terrestrial, Celestial and Orbital Systems	108
	REFERENCES	109
	APPENDIX A: Summary of Reflection and Rotation Matrices	111

LIST OF ILLUSTRATIONS

Figure No.	Title	Page
1-1	Terrestrial, Celestial, and Orbital Coordinate Systems	2
1-2	Types of Coordinate Systems	3
2-1	Polar Motion	11
2-2	Position of Point Moving Uniformly Along Equator Minus Position on Actual Equator	13
2-3	Terrestrial and Geodetic Coordinate Systems	14
2-4	Transformations from Instantaneous to Average Terrestrial System	16
2-5	Reference Ellipsoid	20
2-6	Various Latitudes	23
2-7	Tangent Line to the Meridian Ellipse	24
2-8	Point Above Reference Ellipsoid	29
2-9	Meridian Section of the Earth	34A
2-10	Orientation of Ellipsoid to Geoid	37A
	....	
2-12	Geodetic and Local Geodetic Coordinate Systems	51
2-13	Equations Relating Terrestrial Systems	59
3-1	Ecliptic System	64
3-2	Right Ascension System	66
3-3	Hour Angle System	68
3-4	Time, Longitude, and Right Ascension	70

Figure No.	Title	Page
3-5	Horizon System	72
3-6	Variations of the Celestial Right Ascension System	75
3-7	Motion of Celestial Pole	77
3-8	The Effect of Precession and Nutation	78
3-9	Mean Celestial Coordinate Systems	80
3-10	True and Mean Celestial Coordinate Systems	82
3-11	Celestial Coordinate Systems	91
4-1	Orbit Ellipse	93
4-2	Keplerian Orbital Elements	97
4-3	Satellite Subpoint	100
4-4	Topocentric Coordinates of Satellite	102
5-1	Coordinate Systems	105

LIST OF TABLES

Table No.	Title	Page
2-1	Parameters Defining the 1927 North American Datum	44
2-2	Translation Components	44
2-3	Example of Datum Transformations	49
2-4	Reference Poles, Planes and Axes Defining Terrestrial Coordinate Systems	58
2-5	Transformations Among Terrestrial Coordinate Systems	60
3-1	Reference Poles, Planes and Axes Defining Celestial Coordinate Systems	89
3-2	Transformations Among Celestial Coordinate Systems	90

1. INTRODUCTION

These notes discuss the precise definitions of, and transformations between, the coordinate systems to which coordinates of stations on or above the surface of the earth are referred. To define a coordinate system we must specify:

a) the location of the origin,

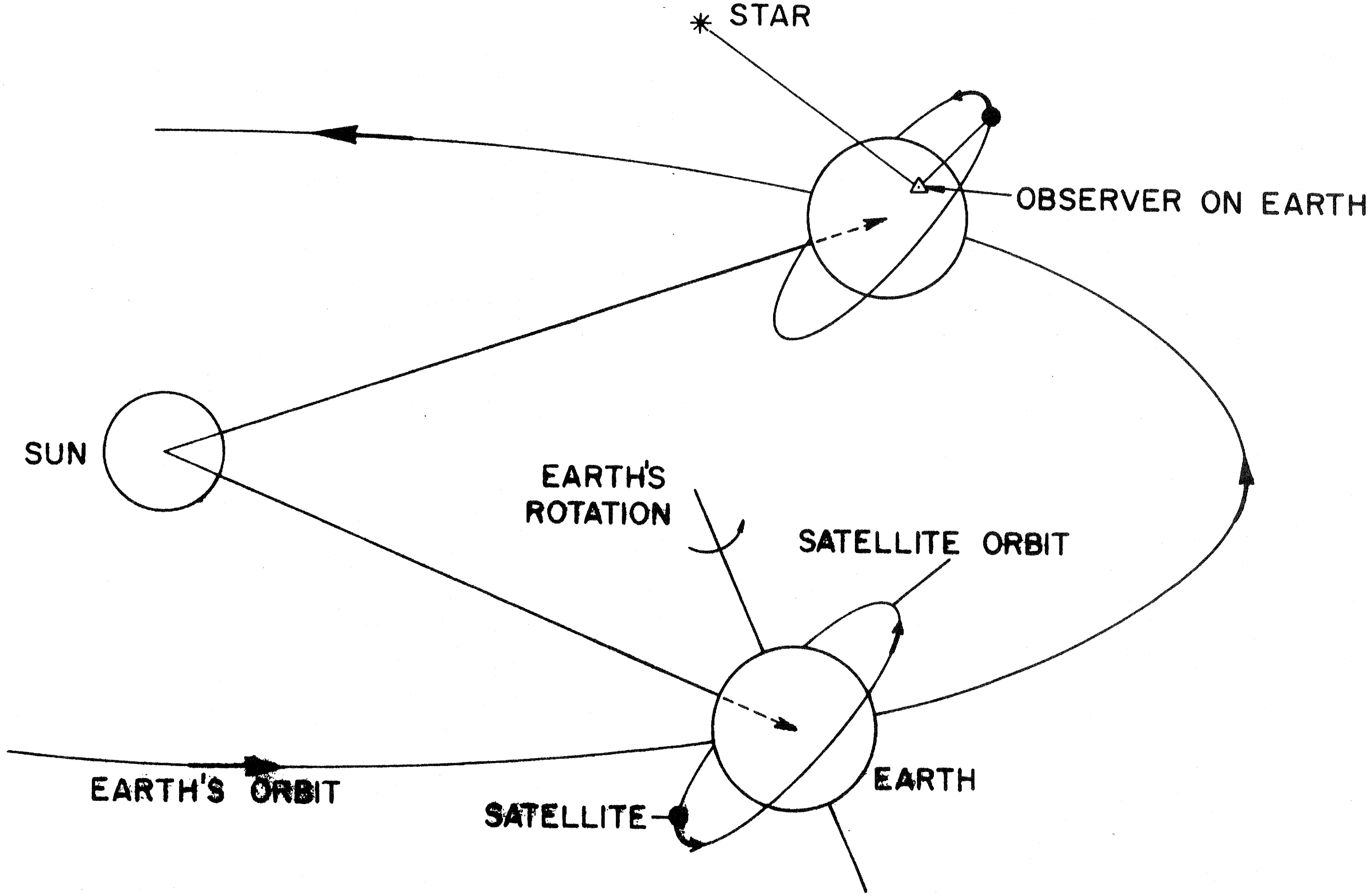
b) the orientation of the three axes,

c) the parameters (Cartesian, curvilinear) which define the position of a point referred to the coordinate system.

The earth has two different periodic motions in space. It <u>rotates</u> about its axis, and it <u>revolves</u> about the sun (see Figure 1-1). There is also one natural satellite (the moon) and many artificial satellites which have a third periodic motion in space: <u>orbital motion</u> about the earth. These periodic motions are fundamental to the definition of systems of coordinates and systems of time.

<u>Terrestrial</u> coordinate systems are earth fixed and rotate with the earth. They are used to define the coordinates of points on the surface of the earth. There are two kinds of terrestrial systems called <u>geocentric</u> systems and <u>topocentric</u> systems (see Figure 1-2).

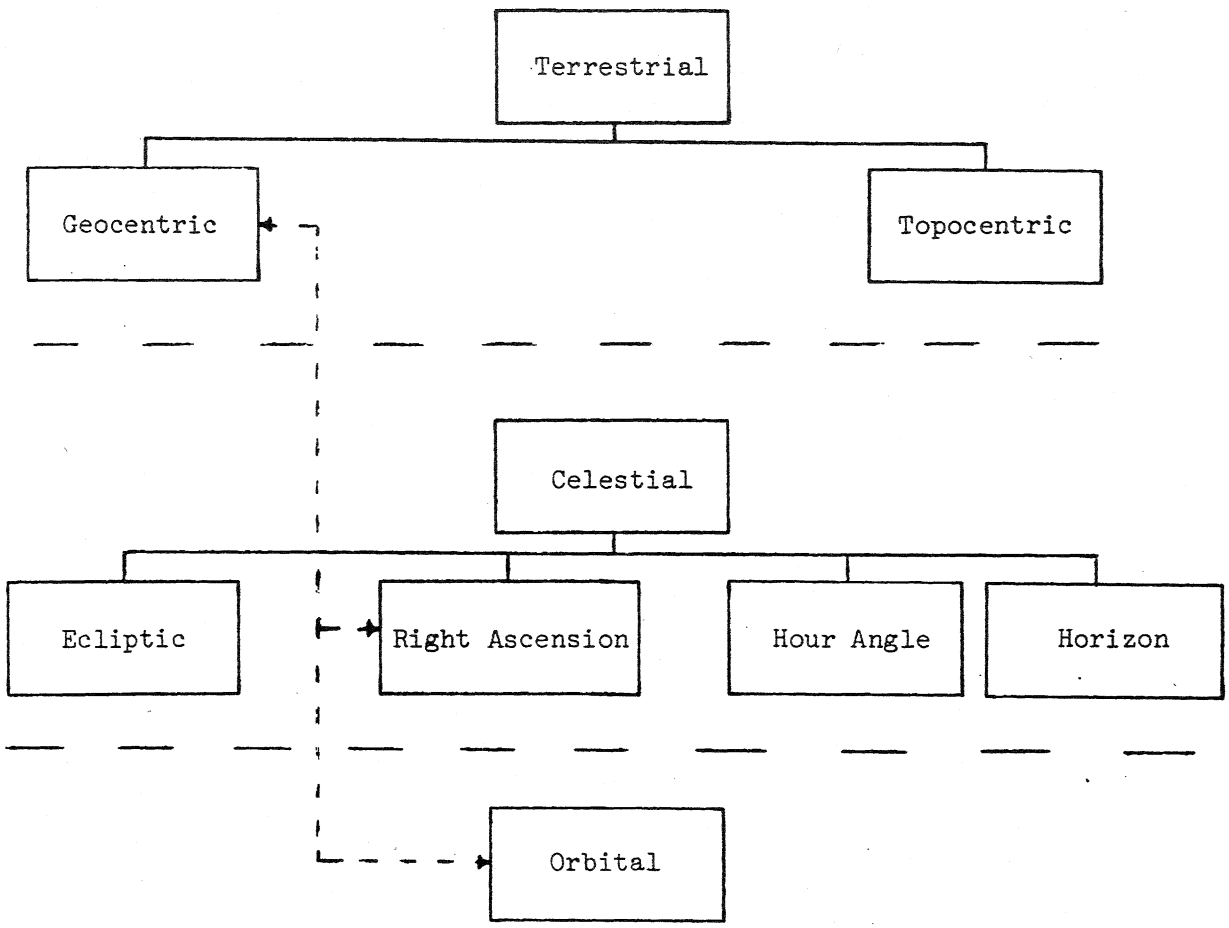
<u>Celestial</u> coordinate systems do not revolve but may rotate with the



TERRESTRIAL , CELESTIAL AND ORBITAL COORDINATE SYSTEMS.

Figure 1-1.

Figure 1-2. Types of Coordinate Systems.



earth. They are used to define the coordinates of celestial bodies such as stars. There are four different celestial systems, called the ecliptic, right ascension, hour angle, and horizon systems.

The orbital system does not rotate with the earth, but revolves with it. It is used to define the coordinates of satellites orbiting around the earth.

1.1 POLES, PLANES AND AXES

The orientation of axes of coordinate systems can be described in terms of primary and secondary poles, primary and secondary planes, and primary, secondary and tertiary axes.

The primary pole is the axis of symmetry of the coordinate system, for example the rotation axis of the earth. The primary plane is the plane perpendicular to the primary pole, for example the earth's equatorial plane. The secondary plane is perpendicular to the primary plane and contains the primary pole. It sometimes must be chosen arbitrarily, for example the Greenwich meridian plane, and sometimes arises naturally, for example the equinoctial plane. The secondary pole is the intersection of the primary and secondary planes. The primary axis is the secondary pole. The tertiary axis is the primary pole. The secondary axis is perpendicular to the other two axes, chosen in the direction which makes the coordinate system either right-handed or left-handed as specified.

We will use either the primary plane or the primary pole, and the primary axis to specify the orientation of each of the coordinate systems named above.

For terrestrial geocentric systems:

a) the origin is near the centre of the earth,

b) the primary pole is aligned to the earth's axis of rotation, and the primary plane perpendicular to this pole is called the equatorial plane,

c) the primary axis is the intersection between the equatorial plane and the plane containing the Greenwich meridian,

d) the systems are right-handed.

For terrestrial topocentric systems:

a) the origin is at a point near the surface of the earth,

b) the primary plane is the plane tangential to the surface of the earth at that point,

c) the primary axis is the north point (the intersection between the tangential plane and the plane containing the earth's north rotational pole),

d) the systems are left-handed.

For the celestial ecliptic system:

a) the origin is near the centre of the sun,

b) the primary plane is the plane of the earth's orbit, called the ecliptic plane,

c) the primary axis is the intersection between the ecliptic plane and the equatorial plane, and is called the vernal equinox,

d) the system is right-handed.

For the celestial right ascension system:

a) the origin is near the centre of the sun,

b) the primary plane is the equatorial plane,

c) the primary axis is the vernal equinox,

d) the system is right-handed.

For the celestial hour angle system:

a) the origin is near the centre of the sun,

b) the primary plane is the equatorial plane,

c) the secondary plane is the celestial meridian (the plane containing the observer and the earth's rotation axis),

d) the system is left-handed.

For the celestial horizon system:

a) the origin is near the centre of the sun,

b) the primary plane is parallel to the tangential plane at the observer (the horizon plane),

c) the primary axis is parallel to the observer's north point,

d) the system is left-handed.

For the orbital system:

a) the origin is the centre of gravity of the earth,

b) the primary plane is the plane of the satellite orbit around the earth,

c) the primary axis is in the orbital plane and is oriented towards the point of perigee (the point at which the satellite most closely approaches the earth) and is called the line of apsides,

d) the system is right-handed.

1.2 UNIVERSAL AND SIDEREAL TIME

Also intimately involved with the earth's periodic rotation and revolution are two systems of time called universal (solar) time (UT) and sidereal time (ST). A time system is defined by specifying an interval and an epoch. The solar day is the interval between successive passages of the sun over the same terrestrial meridian. The sidereal day is the interval between two successive passages of the vernal equinox over the same terrestrial meridian. The sidereal epoch is the angle between the vernal equinox and some terrestrial meridian: if this is the Greenwich meridian then the epoch is Greenwich Sidereal Time (GST). The solar epoch is rigorously related to the sidereal epoch by a mathematical formula. Sidereal time is the parameter relating terrestrial and celestial systems.

1.3 COORDINATE SYSTEMS IN GEODESY

Geodesy is the study of the size and shape of the earth and the determination of coordinates of points on or above the earth's surface.

Coordinates of one station are determined with respect to coordinates of other stations by making one or more of the following four categories of measurements: directions, distances, distance differences, and heights. Horizontal and vertical angular measurements between two stations on the earth (as are measured by theodolite for example) are terrestrial directions. Angular measurements between a station on the earth and a satellite position (as are measured by photographing the satellite in the star background for example) are satellite directions. Angular measurements between a station on the earth and a star (as are measured by direct theodolite pointings on the star for example) are astronomic directions. Distances between two stations on the earth (as are measured by electromagnetic distance measuring instruments for example) are terrestrial distances. Distances between a station on the earth and a satellite position (as are measured by laser ranging for example) are satellite distances. Measurements of the difference in distance between one station on the earth and two other stations (as are measured by hyperbolic positioning systems for example) are terrestrial distance differences. Measurements of the difference in distance between one station on the earth and two satellite positions (as are measured by integrated Doppler shift systems for example) are satellite distance differences. All these measurements determine the geometrical relationship between stations, and are the subject of geometric geodesy [e.g. Bomford 1962].

Spirit level height differences and enroute gravity values are measurements related to potential differences in the earth's gravity field, and are the subject of physical geodesy [e.g. Heiskanen and Moritz 1967].

The functional relationship between these measurements and the coordinates of the stations to and from which they are made is incorporated into a mathematical model. A unique solution for the unknown coordinates can be obtained by applying the least squares estimation process [Wells and Krakiwsky 1971] to the measurements and mathematical model.

Details on coordinate systems as employed for terrestrial and satellite geodesy can be found in Veis [1960] and Kaula [1966], and for geodetic astronomy in Mueller [1969].

2. TERRESTRIAL COORDINATE SYSTEMS

In this chapter we will discuss terrestrial geocentric and terrestrial topocentric coordinate systems.

We first discuss terrestrial geocentric systems using only Cartesian coordinates, and considering in detail what is meant by "the earth's axis of rotation" and "the Greenwich meridian". Then the relationship between Cartesian and curvilinear coordinates is described. Geodetic datums are discussed. Finally terrestrial topocentric systems are considered, with attention paid to what is meant by "the surface of the earth".

2.1 TERRESTRIAL GEOCENTRIC SYSTEMS

In the introduction it was stated that for terrestrial geocentric systems:

a) the origin is near the centre of the earth,

b) the primary pole is aligned to the earth's axis of rotation,

c) the primary axis is the intersection between the primary plane and the plane containing the Greenwich meridian,

d) the systems are right-handed.

The last specification is unambiguous. As we shall see the other three are not. We will first discuss problems in defining the earth's axis of rotation and the Greenwich meridian. Then we will discuss translations of the origin from the centre of the earth.

2.1.1 Polar Motion and Irregular Rotation of the Earth

We think of the earth as rotating about a fixed axis at a uniform rate. In fact, the axis is not fixed and the rate is not uniform.

Over 70 years ago, it was discovered that the direction of the earth's rotation axis moves with respect to the earth's surface. This polar motion is principally due to the fact that the earth's axes of rotation and maximum inertia do not coincide. The resultant motion is irregular but more or less circular and counterclockwise (when viewed from North), with an amplitude of about 5 meters and a main period of 430 days (called the Chandler period).

Two international organizations, the International Polar Motion Service (IPMS) and the Bureau International de l'Heure (BIH) routinely measure this motion through astronomic observations; the IPMS from five stations at the same latitude, and the BIH from about 40 stations scattered worldwide. The results are published as the coordinates of the true rotation axis with respect to a reference point called the Conventional International Origin (CIO) which is the average position of the rotation axis during the years 1900-1905(IUGG (1967) Bull Geod 86, 379 (1967) Resolution 19). Figure 2-1 shows the polar motion during 1969 as determined by IPMS and BIH.

Over 30 years ago irregularities in the rotation of the earth were

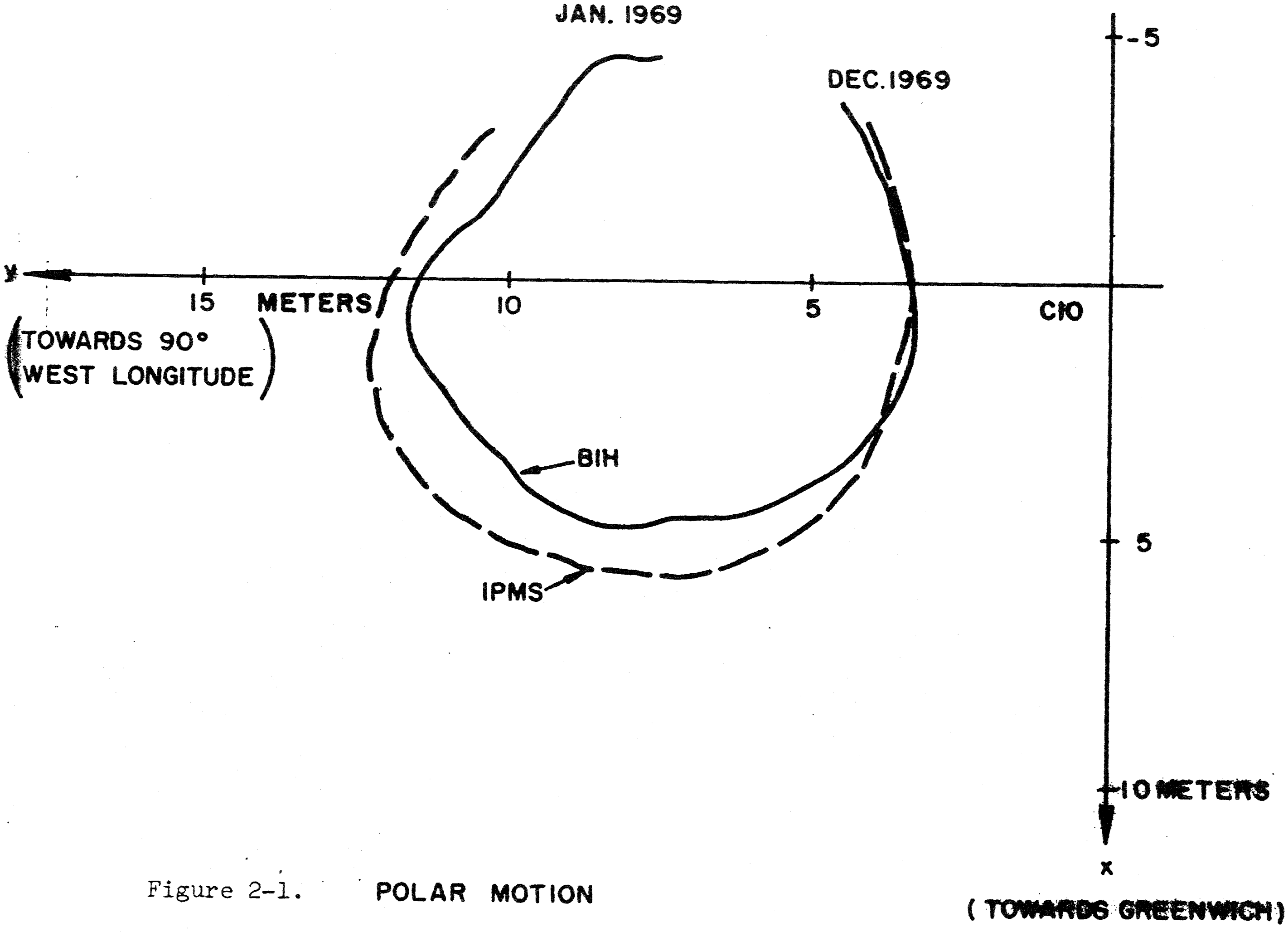


Figure 2-1. POLAR MOTION

discovered (other than polar motion). There are three types of irregularities: seasonal variations probably due to meteorological changes or earth tides; secular decrease due to tidal friction; and irregular fluctuations [Mueller 1969]. The seasonal variation is the only one of these presently being taken into account, and it is more or less reproducible from year to year, and produces a displacement along the equator of up to 15 meters with respect to a point rotating uniformly throughout the year (see Figure 2-2).

Because of this seasonal variation, the Greenwich meridian (the plane containing the earth's rotation axis and the center of the transit instrument at Greenwich Observatory) does not rotate uniformly. A ficticious zero meridian which does rotate uniformly (so far as the effects of polar motion and seasonal variations are concerned) is called the Mean Observatory or Greenwich mean astronomic meridian. Its location is defined by the BIH.

2.1.2 Average and Instantaneous Terrestrial Systems

The average terrestrial (A.T.) system is the ideal world geodetic coordinate system (see Figure 2-3):

a) Its origin is at the centre of gravity of the earth.

b) Its primary pole is directed towards the CIO (the average north pole of 1900-1905), and its primary plane is the plane perpendicular to the primary pole and containing the earth's center of gravity (the average equatorial plane).

c) Its secondary plane is the plane containing the primary

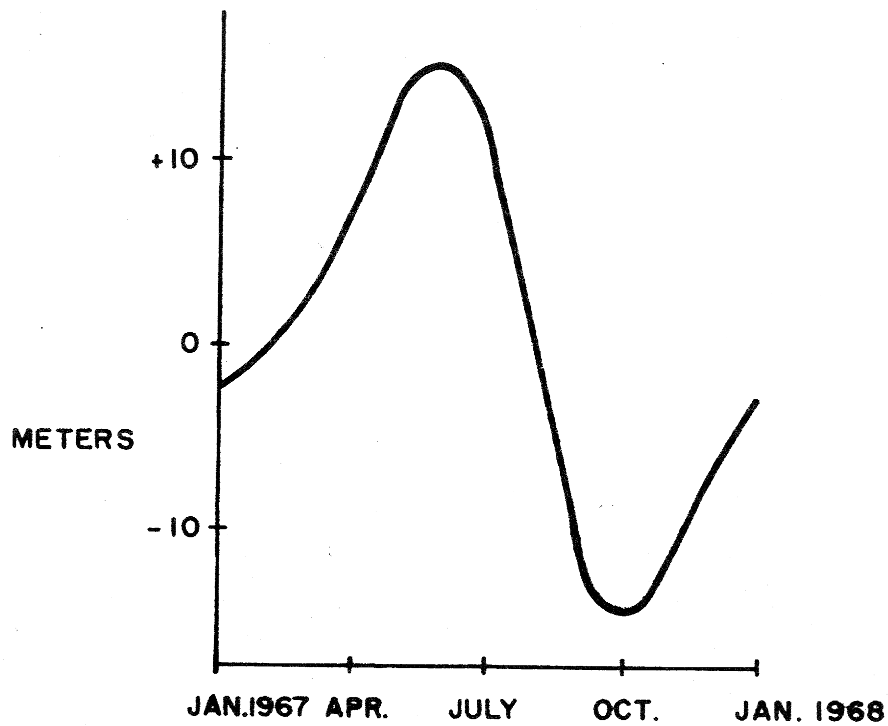


Figure 2-2. POSITION OF POINT MOVING UNIFORMLY ALONG EQUATOR MINUS POSITION OF POINT ON ACTUAL EQUATOR.

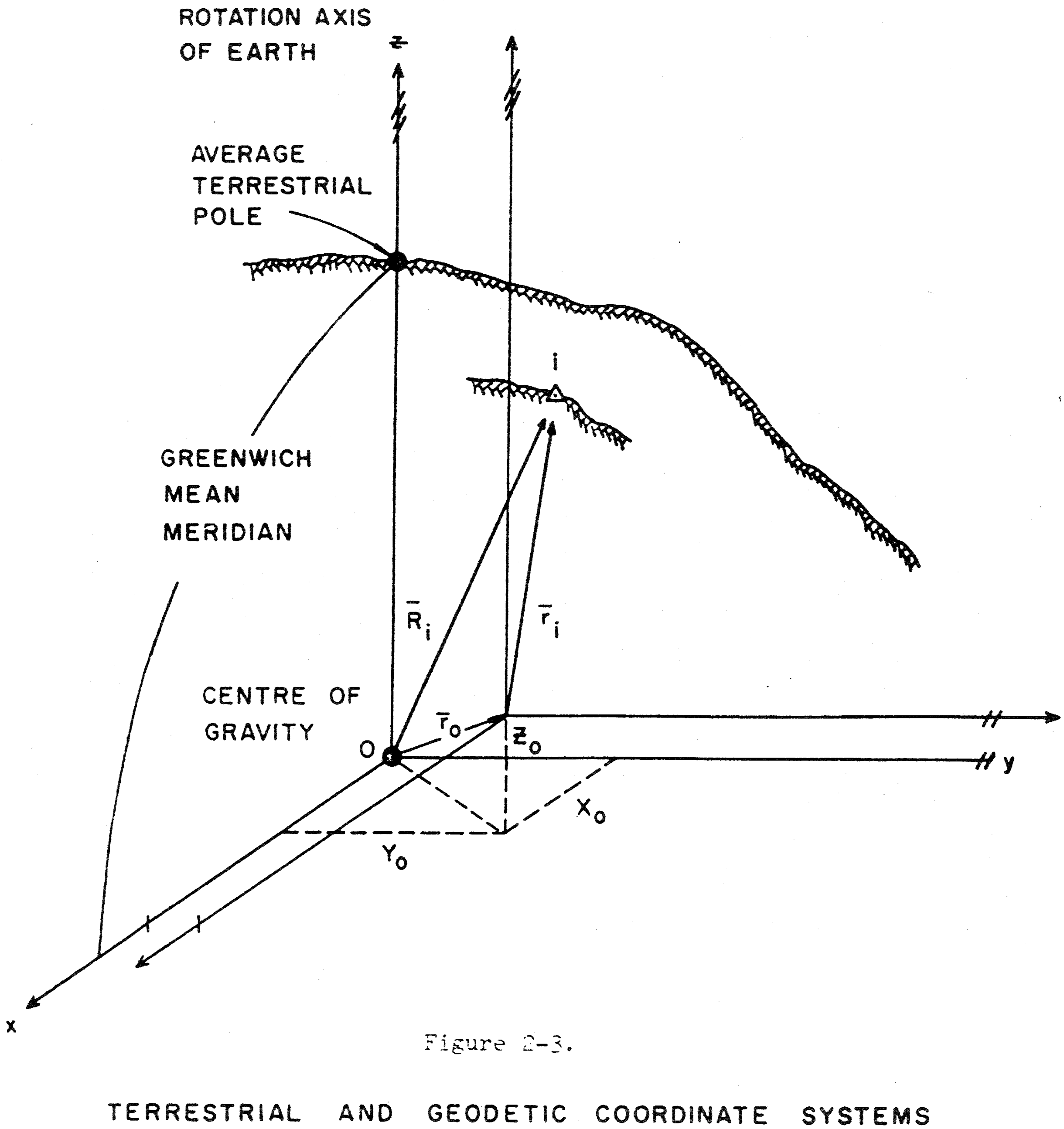


Figure 2-3.

TERRESTRIAL AND GEODETIC COORDINATE SYSTEMS pole and the Mean Observatory. The intersection of these two planes is the secondary pole, or primary axis.

d) It is a right-handed system.

We can then define the position vector $\bar{R}_i$ of a terrain point i in terms of its Cartesian coordinates x, y, z as

$$\bar{R}_i = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} . \qquad 2\text{-}1$$

The instantaneous terrestrial (I.T.) system is specified as follows:

a) Its origin is at the centre of gravity of the earth.

b) Its primary pole is directed towards the true (instantaneous) rotation axis of the earth.

c) Its primary axis is the intersection of the primary plane and the plane containing the true rotation axis and the Mean Observatory.

d) It is a right-handed system.

The main characteristic of these two systems is that they are geocentric systems having their origins at the centre of gravity of the earth and the rotation axis of the earth as the primary pole.

By means of rotation matrices [Thompson 1969; Goldstein 1950; Wells 1971] the coordinates of a point referred to the instantaneous terrestrial system are transformed into the average system by the following equation (see Figure 2-4):

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = R_2(-x_p)\, R_1(-y_p) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{I.T.} ,} \qquad 2\text{-}2$$

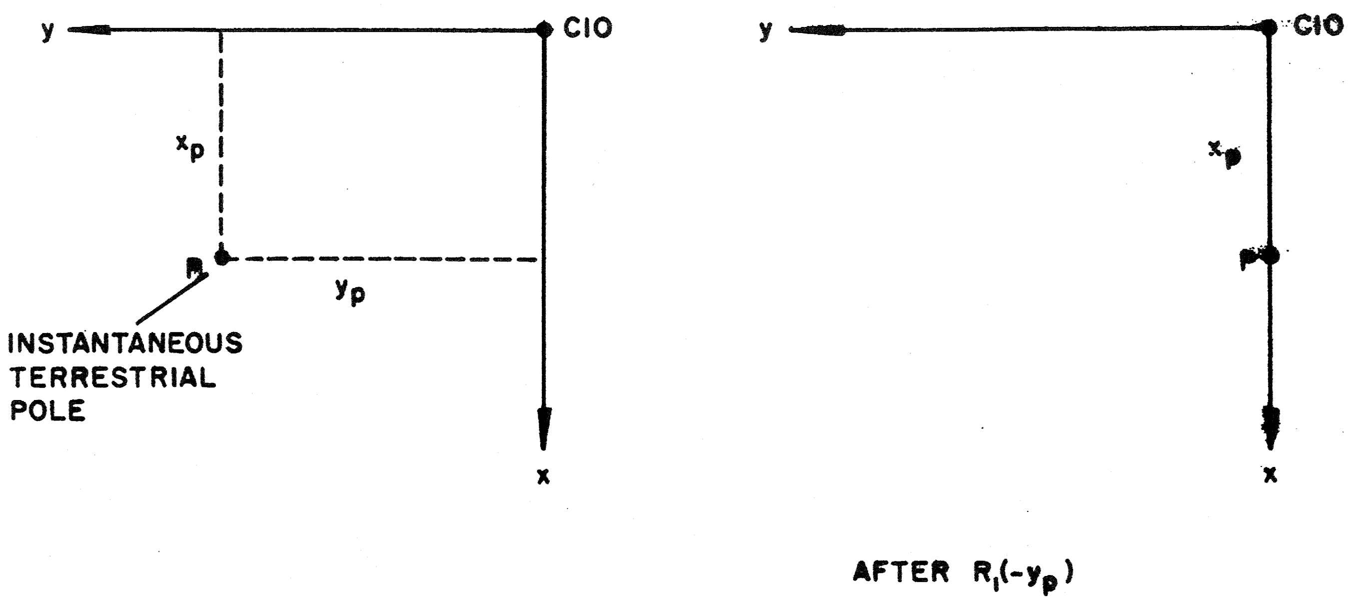


Figure 2-4

TRANSFORMATION FROM INSTANTANEOUS TO AVERAGE TERRESTRIAL SYSTEM.

where (x_p, y_p) are expressed in arcseconds, and the rotation matrices are

$$R_1(-y_p) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(-y_p) & \sin(-y_p) \\ 0 & -\sin(-y_p) & \cos(-y_p) \end{bmatrix} ,$$

a clockwise (negative) rotation about the x axis, and

$$R_2(-x_p) = \begin{bmatrix} \cos(-x_p) & 0 & -\sin(-x_p) \\ 0 & 1 & 0 \\ \sin(-x_p) & 0 & \cos(-x_p) \end{bmatrix} ,$$

a clockwise (negative) rotation about the y-axis. The inverse is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{I.T.} = [R_2(-x_p)\ R_1(-y_p)]^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} ,$$

and because of the orthogonal characteristic of rotation matrices, that is

$$R^{-1}(\theta) = R^T(\theta) = R(-\theta) ,$$

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{I.T.} = R_1(y_p) R_2(x_p) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} .} \qquad 2\text{-}3$$

2.1.3 Geodetic Systems

In terms of Cartesian coordinates, the geodetic (G) coordinate system is that system which is introduced into the earth such that its three axes are coincident with or parallel to the corresponding three axes of the average terrestrial system (see Figure 2-3). The first situation defines a geocentric geodetic system while the latter non-geocentric system is commonly referred to as a relative geodetic system, whose relationship to the average terrestrial system is given by the three datum translation components

$$\bar{r}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} ,$$

and in vector equation form, the relationship is

$$\bar{R}_i = \bar{r}_o + \bar{r}_i ,$$

where the position vector $\bar{r}_i$ is referred to the geodetic system, that is

$$\bar{r}_i = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_G ,$$

and

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} + \begin{bmatrix} x \\ y \\ z \end{bmatrix}_G .} \qquad 2\text{-}4$$

A more detailed account of how a relative geodetic system is established within the earth is in order (Section 2.3), but before this can be done, it would be useful to review the relationship between Cartesian and curvilinear coordinates.

2.2 RELATIONSHIP BETWEEN CARTESIAN AND CURVILINEAR COORDINATES

In this section we first describe the Cartesian (x, y, z) and curvilinear (latitude, longitude, height) coordinates for a point on the reference ellipsoid. We then develop expressions for its position vector in terms of various latitudes. Finally the transformation from geodetic coordinates (ϕ, λ, h) to (x, y, z) and its inverse are discussed.

2.2.1 Cartesian and Curvilinear Coordinates of a Point on the Reference Ellipsoid

The specific ellipsoid used in geodesy as a reference surface is a rotational ellipsoid formed from the rotation of an ellipse about its <u>semi-minor axis</u> b (Figure 2-5). The <u>semi-major axis</u> a and the <u>flattening</u>

$$f = \frac{a - b}{a} \tag{2-5}$$

are the defining parameters of the reference ellipsoid.

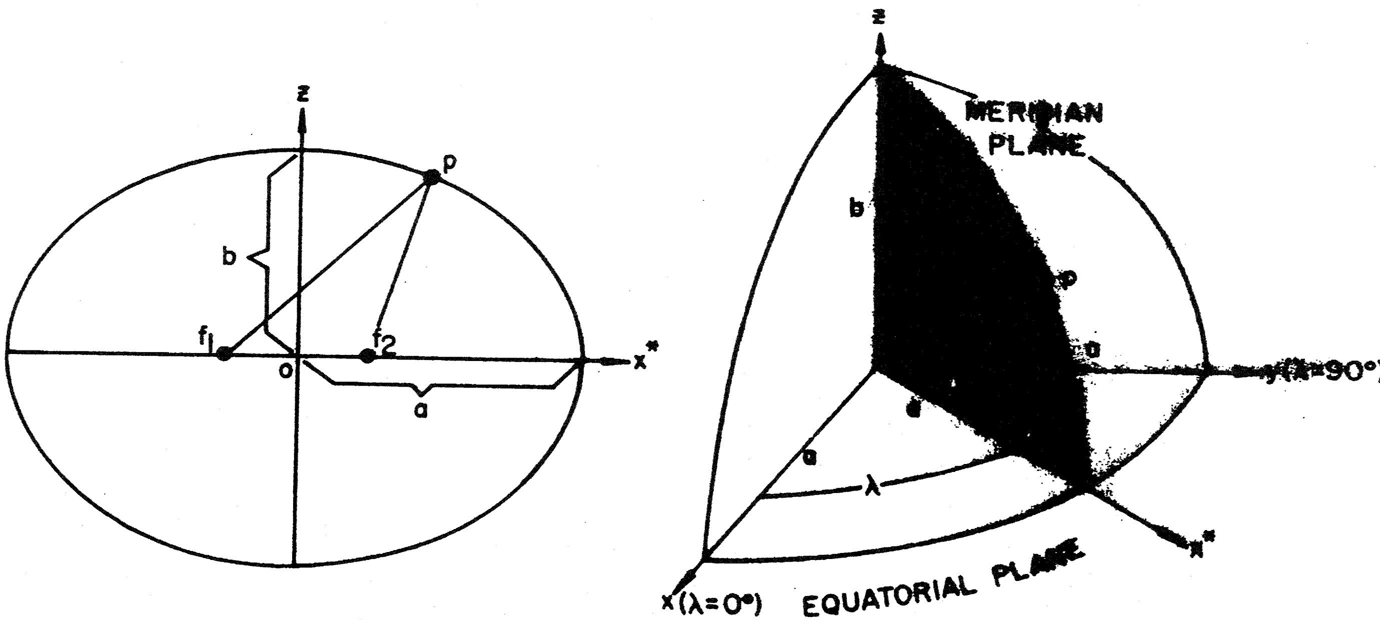
Other useful parameters associated with this particular ellipsoid are the <u>first eccentricity</u>

$$e^2 = \frac{a^2 - b^2}{a^2} \quad , \tag{2-6}$$

and the <u>second eccentricity</u>

$$(e')^2 = \frac{a^2 - b^2}{b^2} \quad . \tag{2-7}$$

A Cartesian coordinate system is superimposed on the reference ellipsoid (see Figure 2-5) so that:



ELLIPSE $f_1p + f_2p = 2a$

Figure 2-5. REFERENCE ELLIPSOID

a) The origin of the Cartesian system is the centre of the ellipsoid.

b) The primary pole (z-axis) of the Cartesian system is the semi-minor axis of the ellipsoid. The primary plane is perpendicular to the primary axis and is called the equatorial plane.

c) Any plane containing the semi-minor axis and cutting the surface of the ellipsoid is called a meridian plane. The particular meridian plane chosen as the secondary plane is called the Greenwich meridian plane. The secondary pole (x-axis) is the intersection of the equatorial plane and the Greenwich meridian plane.

d) The y-axis is chosen to form a right-handed system, and lies in the equatorial plane, 90° counterclockwise from the x-axis.

The equation of this ellipsoid, in terms of Cartesian coordinates is

$$\bar{X}^T S_E \bar{X} = 1 , \qquad 2\text{-}8$$

where

$$\bar{X}^T = [x\ y\ z] ,$$

$$S_E = \begin{bmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/a^2 & 0 \\ 0 & 0 & 1/b^2 \end{bmatrix} , \qquad 2\text{-}9$$

or

$$\boxed{\frac{x^2 + y^2}{a^2} + \frac{z^2}{b^2} = 1 .} \qquad 2\text{-}10$$

The latitude of a point is the acute angular distance between the equatorial plane and the ellipsoid normal through the point measured in the meridian plane of the point. The line perpendicular to the ellipsoid at a point is called the ellipsoid normal at the point. Ellipsoid normals only pass through the geometric center of the ellipsoid in the equatorial plane or along the semi-minor axis. Therefore there are two different kinds of latitude. The angle between the ellipsoid normal at the point and the equatorial plane is called the geodetic latitude ϕ. The angle between the line joining the point to the centre of the ellipse, and the equatorial plane is called the geocentric latitude ψ. There is also a third latitude, used mostly as a mathematical convenience, called the reduced latitude β (see Figure 2-6).

The longitude λ of a meridian plane is the counterclockwise angular distance between the Greenwich meridian plane and the meridian plane of the point, measured in the equatorial plane (see Figure 2-5).

The ellipsoid height h of a point is its linear distance above the ellipsoid, measured along the ellipsoidal normal at the point (see Figure 2-8).

2.2.2 The Position Vector in Terms of the Geodetic Latitude

Consider a point P on the surface of the ellipsoid. The coordinates of P referred to a system with the primary axis (denoted x^*) in the meridian plane of P are

$$\bar{r} = \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix} . \qquad 2\text{-}11$$

The plane perpendicular to the ellipsoid normal at P, and passing through P is called the tangent plane at P. From Figure 2-7 the slope of the tangent plane is

$$\frac{dz}{dx^*} = \tan\,(90^\circ+\phi) = -\,\frac{\cos\phi}{\sin\phi} . \qquad 2\text{-}12$$

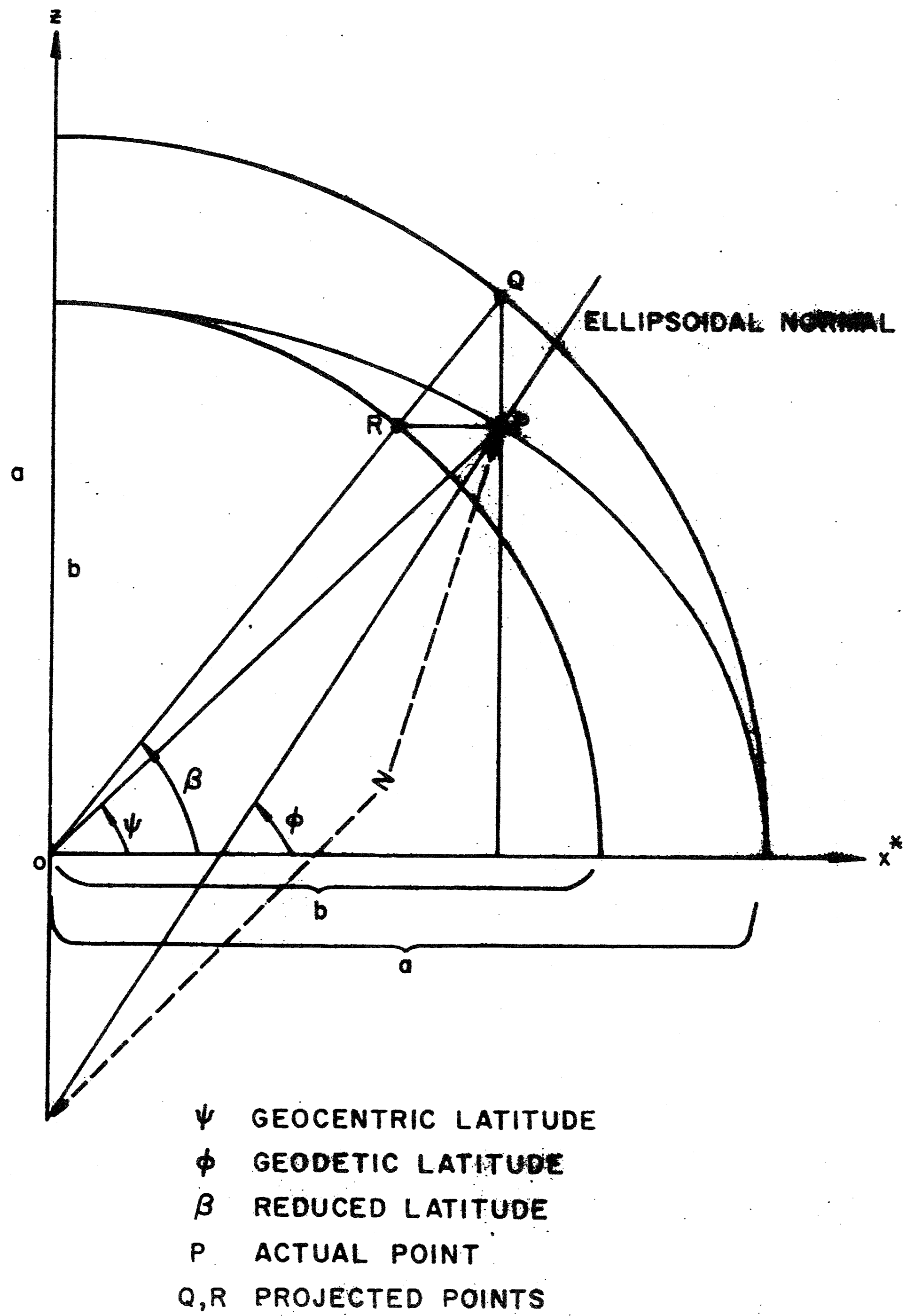


Figure 2-6. VARIOUS LATITUDES

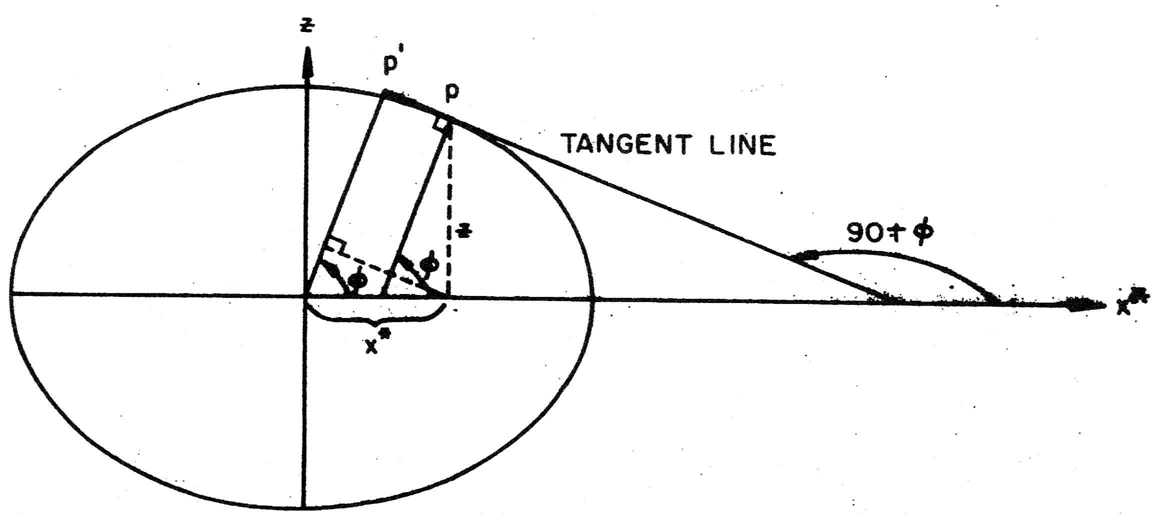


Figure 2-7. TANGENT LINE TO THE MERIDIAN ELLIPSE

The slope can also be computed from the equation of the meridian ellipse as follows:

$$\frac{(x^*)^2}{a^2} + \frac{z^2}{b^2} = 1 \qquad 2\text{-}13$$

or

$$b^2\,(x^*)^2 + a^2\,z^2 = a^2\,b^2 \qquad 2\text{-}14$$

$$2b^2 x^*\,dx^* + 2a^2\,zdz = 0 \qquad 2\text{-}15$$

$$\frac{dz}{dx^*} = -\frac{b^2\,x^*}{a^2\,z} \quad . \qquad 2\text{-}16$$

It follows from the above two equations for the slope, that

$$\frac{b^2\,x^*}{a^2\,z} = \frac{\cos\phi}{\sin\phi} \qquad 2\text{-}17$$

or

$$b^2\,(x^*)\,\sin\phi = a^2\,z\,\cos\phi \qquad 2\text{-}18$$

and after squaring the above

$$b^4\,(x^*)^2\,\sin^2\phi = a^4\,z^2\,\cos^2\phi \quad . \qquad 2\text{-}19$$

Expressing equations 2-14 and 2-19 in matrix form

$$\begin{bmatrix} b^4\,\sin^2\phi & -a^4\,\cos^2\phi \\ b^2 & a^2 \end{bmatrix} \begin{bmatrix} (x^*)^2 \\ z^2 \end{bmatrix} = \begin{bmatrix} 0 \\ a^2 b^2 \end{bmatrix} . \qquad 2\text{-}20$$

The inverse of the coefficient matrix is

$$\frac{1}{a^2 b^2 (a^2\,\cos^2\phi + b^2\,\sin^2\phi)} \begin{bmatrix} a^2 & a^4\,\cos^2\phi \\ -b^2 & b^4\,\sin^2\phi \end{bmatrix} ,$$

therefore

$$\begin{bmatrix} (x^*)^2 \\ z^2 \end{bmatrix} = \frac{1}{a^2\,\cos^2\phi + b^2\,\sin^2\phi} \begin{bmatrix} a^4\,\cos^2\phi \\ b^4\,\sin^2\phi \end{bmatrix}$$

and finding the square root

$$\begin{bmatrix} x^* \\ z \end{bmatrix} = \frac{1}{(a^2\,\cos^2\phi + b^2\,\sin^2\phi)^{1/2}} \begin{bmatrix} a^2\,\cos\phi \\ b^2\,\sin\phi \end{bmatrix} . \qquad 2\text{-}21$$

From Figure 2-6

$$\cos\phi = \frac{x^*}{N} \quad ,$$

but from equation 2-21

$$x^* = \frac{a^2 \cos\phi}{(a^2 \cos^2\phi + b^2 \sin^2\phi)^{1/2}} \quad ,$$

therefore

$$\boxed{N = \frac{a^2}{(a^2\cos^2\phi + b^2\sin^2\phi)^{1/2}}} \qquad 2\text{-}22$$

$$\boxed{\bar{r} = \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} N\cos\phi \\ 0 \\ N\, b^2/a^2 \sin\phi \end{bmatrix}} \quad . \qquad 2\text{-}23$$

N is the radius of curvature of the ellipsoid surface in the plane perpendicular to the meridian plane (called the <u>prime vertical</u> plane).

We now refer the position vector P to a system with the primary axis in the Greenwich meridian plane, that is we rotate the coordinate system about the z-axis clockwise (negative rotation) through the longitude λ.

$$\bar{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = R_3(-\lambda) \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix}$$

$$= \begin{bmatrix} \cos(-\lambda) & \sin(-\lambda) & 0 \\ -\sin(-\lambda) & \cos(-\lambda) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} N\cos\phi \\ 0 \\ N\, b^2/a^2 \sin\phi \end{bmatrix}$$

or

$$\boxed{\bar{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = N \begin{bmatrix} \cos\phi & \cos\lambda \\ \cos\phi & \sin\lambda \\ b^2/a^2 & \sin\phi \end{bmatrix}.} \tag{2-24}$$

2.2.3 The Position Vector in Terms of the Geocentric and Reduced Latitudes

From Figure 2-6 the position vector of the point P in terms of the geocentric latitude ψ is

$$\bar{r} = \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix} = |\bar{r}| \begin{bmatrix} \cos\psi \\ 0 \\ \sin\psi \end{bmatrix}$$

where $|\bar{r}|$ is the magnitude of $\bar{r}$.

Rotating the coordinate system to introduce longitude as before,

$$\boxed{\bar{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = R_3(-\lambda) \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix} = |\bar{r}| \begin{bmatrix} \cos\psi & \cos\lambda \\ \cos\psi & \sin\lambda \\ \multicolumn{2}{c}{\sin\psi} \end{bmatrix}.} \tag{2-25}$$

From Figure 2-6 the reduced latitude β of the point P is the geocentric latitude of both the points Q and R, where Q is the projection of P parallel to the semi-minor axis to intersect a circle with radius equal to the semi-major axis, and R is the projection of the point P parallel to the semi-major axis to intersect a circle with radius equal to the semi-minor axis.

The position vector of P in terms of the reduced latitude β is

$$\bar{r} = \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} a\cos\beta \\ 0 \\ b\sin\beta \end{bmatrix}.$$

Rotating the coordinate system to introduce longitude,

$$\boxed{\bar{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = R_3(-\lambda) \begin{bmatrix} x^* \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} a \cos\beta \quad \cos\lambda \\ a \cos\beta \quad \sin\lambda \\ b \sin\beta \end{bmatrix} .} \qquad 2\text{-}26$$

2.2.4 Relationships between Geodetic, Geocentric and Reduced Latitudes

From equations 2-24, 2-25, and 2-26

$$\frac{z}{x} = \frac{b^2}{a^2} \tan\phi \cos\lambda = \tan\psi \cos\lambda = \frac{b}{a} \tan\beta \cos\lambda \ .$$

Cancelling the cos λ term,

$$\boxed{\tan\beta = \frac{b}{a} \tan\phi \quad ,} \qquad 2\text{-}27$$

$$\boxed{\tan\beta = \frac{a}{b} \tan\psi \quad ,} \qquad 2\text{-}28$$

$$\boxed{\tan\psi = \frac{b^2}{a^2} \tan\phi \quad .} \qquad 2\text{-}29$$

2.2.5 The Position Vector of a Point Above the Reference Ellipsoid

Let us consider a terrain point i, as depicted in Figure 2-8, whose coordinates are the geodetic latitude ϕ and longitude λ, and the ellipsoid height h. The projection of i onto the surface of the ellipsoid is along the ellipsoidal normal defined by the unit vector $\hat{u}_z$.

The position vector of i is then the sum of two vectors, namely

$$\bar{r}_i = \bar{r}_p + h\,\hat{u}_z \quad , \qquad 2\text{-}30$$

where $\bar{r}_p$ is defined by equation 2-24 and $\hat{u}_z$ is the unit vector defined by equation 2-68c, that is

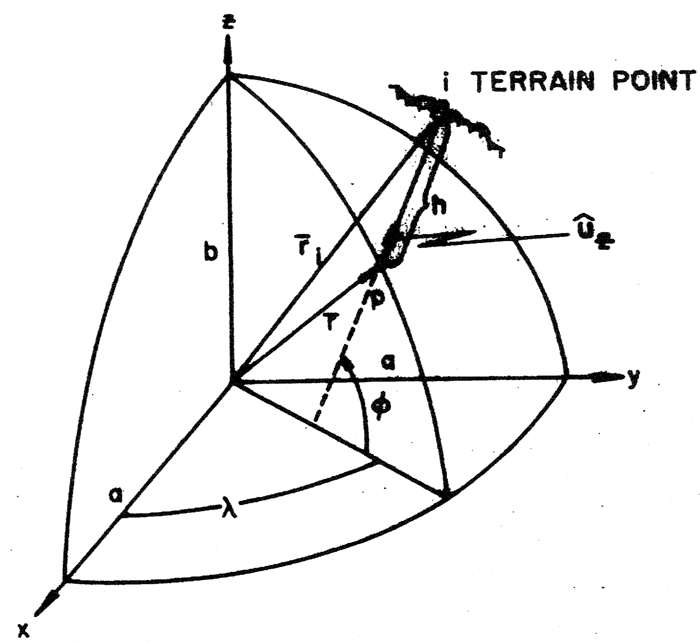


Figure 2-8. POINT ABOVE REFERENCE ELLIPSOID

$$\hat{u}_z = \begin{bmatrix} \cos\phi \ \ \cos\lambda \\ \cos\phi \ \ \sin\lambda \\ \sin\phi \end{bmatrix} .$$

Thus

$$\bar{r}_i = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = N \begin{bmatrix} \cos\phi \ \ \cos\lambda \\ \cos\phi \ \ \sin\lambda \\ b^2/a^2 \ \sin\phi \end{bmatrix} + h \begin{bmatrix} \cos\phi \ \ \cos\lambda \\ \cos\phi \ \ \sin\lambda \\ \sin\phi \end{bmatrix}$$

or

$$\boxed{\bar{r}_i = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (N+h) \ \cos\phi \ \cos\lambda \\ (N+h) \ \cos\phi \ \sin\lambda \\ (Nb^2/a^2+h) \ \sin\phi \end{bmatrix} .} \qquad 2\text{-}31$$

Now the position vector $\bar{r}_i$ in equation 2-31 refers to a coordinate system whose origin is at the geometrical centre of the ellipsoid. If this ellipsoid defines a relative geodetic system, then its centre will not in general coincide with the centre of gravity of the earth. The expression for the position vector in the average terrestrial system is, from equation 2-4

$$(\bar{r}_i)_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} + (\bar{r}_i)_G$$

or

$$\boxed{(\bar{r}_i)_{A.T.} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} + \begin{bmatrix} (N+h) \ \cos\phi \ \cos\lambda \\ (N+h) \ \cos\phi \ \sin\lambda \\ (Nb^2/a^2+h) \ \sin\phi \end{bmatrix} .} \qquad 2\text{-}32$$

This expression gives the general transformation from relative geodetic coordinates (ϕ, λ, h) to average terrestrial coordinates (x, y, z), given the size of the ellipsoid (a, b) and the translation components (x_o, y_o, z_o).

2.2.6 Transformation from Average Terrestrial Cartesian to Geodetic Coordinates

A very useful transformation is the inverse of equation 2-32. Given the average terrestrial coordinates (x, y, z), the translation components (x_o, y_o, z_o), and the size of the ellipsoid (a, b), compute the relative geodetic coordinates (ϕ, λ, h).

First we translate the origin from the centre of gravity to the centre of the ellipsoid. From equation 2-32

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_G = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} - \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} . \qquad 2\text{-}33$$

The longitude λ is computed directly from

$$\boxed{\lambda = \tan^{-1} \left(\frac{y}{x}\right) .} \qquad 2\text{-}34$$

The latitude ϕ and ellipsoid height h are more difficult to compute since N is a function of ϕ, from equation 2-22

$$\boxed{N = \frac{a}{(\cos^2\phi + b^2/a^2 \sin^2\phi)^{1/2}} ,} \qquad 2\text{-}35$$

and h is not known. We begin by computing

$$e^2 = 1 - b^2/a^2 \qquad 2\text{-}36$$

$$p = (x^2 + y^2)^{1/2} . \qquad 2\text{-}37$$

From equation 2-31

$$p^2 = (N+h)^2 \cos^2\phi \cos^2\lambda + (N+h)^2 \cos^2\phi \sin^2\lambda$$

$$p = (N+h) \cos\phi$$

or

$$\boxed{h = \frac{p}{\cos\phi} - N .} \qquad 2\text{-}38$$

Also from 2-31

$$z = (N\, b^2/a^2 + h) \sin\phi$$

$$= (N - \frac{a^2-b^2}{a^2} N + h) \sin\phi$$

$$= (N + h - e^2 N) \sin\phi .$$

Therefore

$$\frac{z}{p} = \frac{(N + h - e^2N) \sin\phi}{(N + h) \cos\phi} = \tan\phi \, (1 - \frac{e^2N}{N+h}) \qquad 2\text{-}39$$

This equation can be developed in two ways, to produce either a direct solution for ϕ which is quite involved, or an iterative solution which is simpler. We consider the iterative solution first. We have

$$\boxed{\phi = \tan^{-1}\left[(\frac{z}{p}) \; (1 - \frac{e^2N}{N+h})^{-1} \right] .}$$

The iterative procedure is initiated by setting

$$\boxed{\begin{aligned} N_o &= a \\ h_o &= (x^2+y^2+z^2)^{1/2} - (ab)^{1/2} \\ \phi_o &= \tan^{-1} [(\frac{z}{p}) \; (1- \frac{e^2N_o}{N_o+h_o})^{-1}] \quad . \end{aligned}}$$

Each iteration then consists of evaluating in order

$$\boxed{\begin{aligned} N_i &= \frac{a}{(\cos^2\phi_{i-1} + b^2/a^2 \sin^2\phi_{i-1})^{1/2}} \\ h_i &= \frac{p}{\cos\phi_{i-1}} - N_i \\ \phi_i &= \tan^{-1} \left[(\frac{z}{p}) \; (1- \frac{e^2N_i}{N_i+h_i})^{-1} \right] . \end{aligned}}$$

The iterations are repeated until

$$(h_i - h_{i-1}) < a\varepsilon$$

and

$$(\phi_i - \phi_{i-1}) < \varepsilon$$

for some appropriately chosen value of ε (for example $\varepsilon = 10^{-10}$ for double precision Fortran on the IBM 360 computer).

Returning to equation 2-39, we eliminate h using equation 2-38 to obtain

$$\frac{z}{p} = \tan\phi(1 - \frac{e^2 N \cos\phi}{p})$$

or

$$p \tan\phi - z = e^2 N \sin\phi \ .$$

In this equation the only unknown is ϕ. We will now modify this equation to obtain an equation which can be solved for $\tan\phi$. Substituting the expression for N from equation 2-35 we have

$$p \tan\phi - z = \frac{a\, e^2 \sin\phi}{(\cos^2\phi + b^2/a^2 \sin^2\phi)^{1/2}} \quad .$$

Dividing the numerator and denominator of the right hand side by cos

$$p \tan\phi - z = \frac{a\, e^2 \tan\phi}{(1 + b^2/a^2 \tan^2\phi)^{1/2}}$$

or

$$(p \tan\phi - z)\ (1 + (1 - e^2) \tan^2\phi)^{1/2} = ae^2 \tan\phi$$

Squaring this equation to eliminate the square root

$$(p^2 \tan^2\phi - 2\, p\, z \tan\phi + z^2)\ (1 + (1 - e^2) \tan^2\phi)$$
$$= a^2 e^4 \tan^2\phi$$

or

$$p^2 \tan^4\phi - 2\, p\, z \tan^3\phi + (\beta + z^2) \tan^2\phi - \frac{2\, p\, z \tan\phi}{(1 - e^2)} + \frac{z^2}{(1 - e^2)} = 0$$

where

$$\beta = \frac{p^2 - a^2 e^4}{(1 - e^2)} \quad .$$

This is a quartic (biquadratic) equation in $\tan\phi$, in which the values of all coefficients are known. Standard procedures for solving quartic equations exist (see for example Korn and Korn, 1968), and have been applied to this equation by Paul (1973), to produce a computer program which is about 25% faster than iterative programs. Once a solution for $\tan\phi$ is obtained, N and h are computed from equations 2-35 and 2-38 respectively.

2.3 GEODETIC DATUMS

There are two natural figures of the earth (see Figure 2-9); the topographic or physical surface of the earth including the surface of the oceans (the <u>terrain</u>), and the equipotential surface of the earth's gravity field which coincides with an idealized surface of the oceans (the <u>geoid</u>).

Control measurements (e.g. distances, angles, spirit levelling) are made between points on the terrain which we call <u>control points</u>. These measurements are used to determine the geometrical relationship between the control points in a computation called <u>network adjustment</u>. Other points are then related to the network of control points through further measurements and computations called <u>densification</u>. The classical approach is to treat the vertical measurements, networks and computations separately from the horizontal measurements, networks and computations. However the unified three dimensional approach is currently gaining favour. [Hotine, 1969].

In the classical approach vertical measurements and networks are referred to a coordinate surface or (vertical) datum which is the geoid. Rather than using the geoid as the coordinate surface or datum for the horizontal measurements and networks as well, a third, unnatural figure of

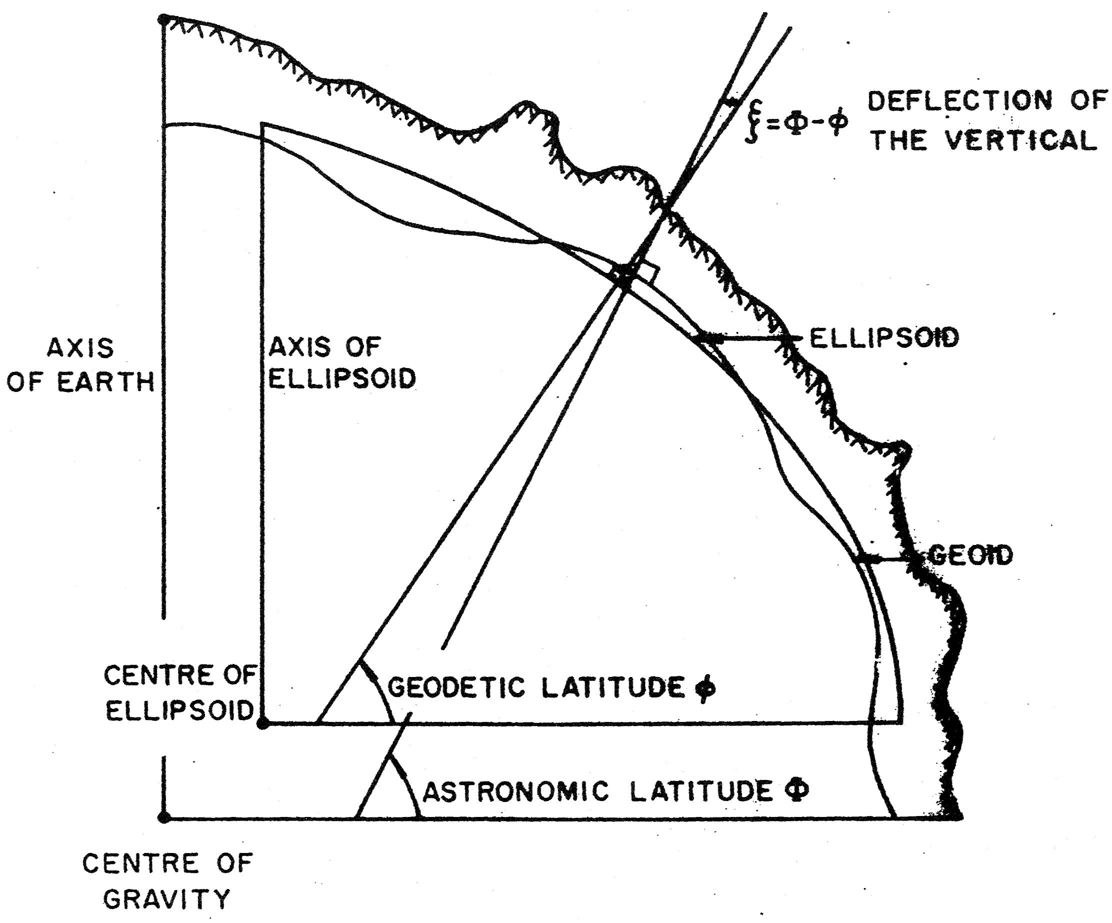


Figure 2-9. **MERIDIAN SECTION OF THE EARTH**

the earth is introduced - the ellipsoid of rotation discussed earlier. The reason a mathematical figure like the ellipsoid is used as the horizontal datum is to simplify the computations required both for network adjustment and densification.

Correction terms are necessary in these computation to account for the fact that the datum is not the geoid. An ellipsoid can be chosen to approximate the geoid closely enough that these correction terms can be assumed linear and for some applications even ignored. For a well-chosen ellipsoid (see Figure 2-9), the geoid-ellipsoid separation (geoid height) is always less than 100 metres, and the difference between the geoid and ellipsoid normals at any point (deflection of the vertical) is usually less than 5 arcseconds, very rarely exceeding 1 arcminute.

Even simpler surfaces than the ellipsoid (such as the sphere or the plane) can be sufficient approximations to the geoid if the area under consideration is sufficiently small, and/or the control application permits lower orders of accuracy.

The introduction of a new surface (the ellipsoid) has a price. The horizontal control network (that is the coordinates of the points of the network) is to be referred to the ellipsoid. Therefore before network computations can begin, the control measurements must first be reduced so that they too "refer" to the ellipsoid.

It is important to distinguish between the datum (the coordinate surface or ellipsoid surface) and the coordinates of the points of the network referred to the datum. It is a common but confusing practice (particularly in North America) to use the term "datum" for the set of coordinates.

2.3.1 Datum Position Parameters

In order to establish an ellipsoid as the reference surface for a system of control we must specify its size and shape (usually by assigning values to the semi-major axis and flattening) and we must specify its position with respect to the earth. A well-positioned ellipsoid will closely approximate the geoid over the area covered by the network for which it is in datum. The parameters to which we assign values in order to specify the ellipsoid position we call the datum position parameters.

In three-dimensional space, any figure (and particularly our ellipsoid) has six degrees of freedom, that is six ways in which its position with respect to a fixed figure (in our case the earth) can be changed. Thus there are six datum position parameters.

Another way of looking at this is to consider two three-dimensional Cartesian coordinate systems, one fixed to the ellipsoid and one fixed to the earth. In general the origins of the two systems will not coincide, and the axes will not be parallel. Therefore, to define the transformation from one system to the other we must specify the location of one origin with respect to the other system, and the orientation of one set of axes with respect to the other system, that is three coordinates, and three rotation angles. These six parameters provide a description of the six degrees of freedom and assigning values to them positions the ellipsoid with respect to the earth. They are our six datum position parameters. A datum then is completely specified by assigning values to eight parameters - the ellipsoid size and shape, and the six datum position parameters.

There are in fact two kinds of datum position parameters in use. One kind is obtained by considering the ellipsoid-fixed and earth-fixed coordinate systems to have their origins in the neighbourhood of the geocentre. The other kind is obtained by considering the ellipsoid-fixed and earth-fixed coordinate systems to have their origins near the surface of the earth at a point we call the <u>initial point</u> of the datum.

In the first (geocentric) case the earth-fixed system is the Average Terrestrial system of section 2.1.2, and the ellipsoid-fixed system is the geodetic system of Equation 2-31 (except that here we assume the geodetic and average terrestrial axes are <u>not</u> in general parallel). In this case the datum position parameters are the Average Terrestrial coordinates of the ellipsoid origin (x_o, y_o, z_o of Equation 2-32) and three rotation angles (say w_1, w_2, w_3) required to define the misalignment between the axes. It is of course highly desirable that the ellipsoid be positioned so that these angles are as small as possible, particularly that the two axes of symmetry (the ellipsoid minor axis and earth's average rotation axis or Average Terrestrial z-axis) be parallel.

In the second (topocentric) case the earth-fixed system is a local astronomic system at the initial point, and the ellipsoid-fixed system is a local geodetic system at the same point (local astronomic and geodetic systems are discussed in section 2.4).

Before proceeding further let us consider the geometry in the neighbourhood of a point on the earth's surface. Figure 2-10 is an exaggerated view of the geodetic meridian plane at such a point, showing the sectioned ellipsoid, geoid, several equipotential surfaces related to the geoid, and the terrain. A particular ellipsoid normal intersects the

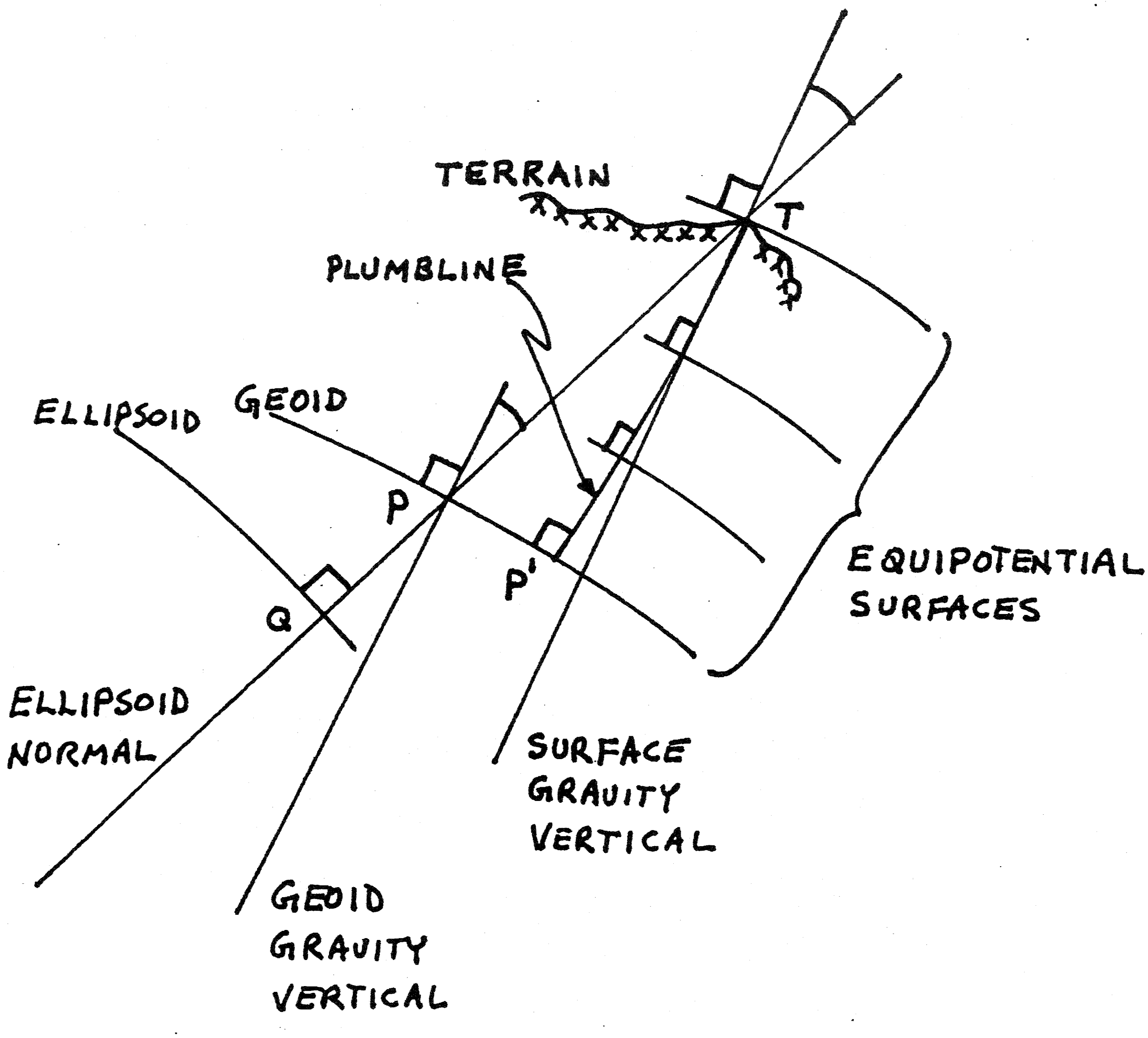


FIGURE 2-10

ORIENTATION OF ELLIPSOID TO GEOID

ellipsoid, geoid, and terrain at Q,P and T respectively. There are three "natural" normals corresponding to this ellipsoid normal; the surface gravity vertical (perpendicular to the equipotential surface at T, passing through T), the geoid gravity vertical (perpendicular to the geoid passing through P), and the plumbline (perpendicular to all equipotential surfaces between terrain and geoid, passing through T). In general, the plumbline is curved while the others are straight lines, and none of these three actually lie in the geodetic meridian plane - they are shown here as projections onto this plane. If the curvature of the plumbline is ignored the two gravity verticals are parallel.

The astronomic meridian plane is the plane containing one of the gravity verticals and a parallel to the Average Terrestrial z-axis. The angle between the gravity vertical and the parallel to the A.T. z-axis is the astronomic co-latitude $(\frac{\pi}{2} - \Phi)$. The angle between the astronomic meridian plane and a reference meridian plane (Greenwich) is the astronomic longitude Λ. The angle between the ellipsoid normal and the gravity vertical is the deflection of the vertical, which can be resolved into a component ξ in the geodetic meridian plane and a component η in the geodetic prime vertical plane (the plane perpendicular to the geodetic meridian plane which contains the ellipsoid normal). Thus corresponding to the two gravity verticals, there are two sets of values for the astronomic latitude and longitude and deflection components, and if the curvature of the plumbline is ignored, these two sets are equal.

If the ellipsoid is positioned so that its geocentric axes are paralled to the Average Terrestrial axes (that is $w_1 = w_2 = w_3 = 0$) then

$$\xi = \Phi - \phi \qquad 2\text{-}40$$

$$\eta = (\Lambda - \lambda) \cos \phi \qquad 2\text{-}41$$

where (ϕ, λ) are the common geodetic coordinates of Q, P and T.

The distance between the ellipsoid and geoid, measured along the ellipsoid normal (QP) is the geoid height N^*. The distance between the ellipsoid and terrain, measured along the ellipsoid normal (QT) is the ellipsoid height h. The distance between the geoid and terrain, measured along the plumbline (P'T) is the orthometric height H. If the curvature of the plumbline is ignored

$$h = N^* + H. \tag{2-42}$$

Given a point some distance from T, the angle between the geodetic meridian plane and the plane containing this point and the ellipsoid normal QPT is the geodetic azimuth α of that point with respect to Q, P or T. (Actually this is the azimuth of the normal section, and is related to the geodetic azimuth by small corrections (Bomford, 1971)). The angle between the astronomic meridian plane and the plane containing this point and the corresponding gravity vertical is the astronomic azimuth A of that point with respect to either P or T depending on which gravity vertical is used. Because the deflection of the vertical is small, then for all such points the difference

$$\delta\alpha = A - \alpha \tag{2-43}$$

is nearly constant, and is the angle between the geodetic and astronomic meridian planes.

Returning to the topocentric datum position parameters, it is natural to specify that our local geodetic system at the initial point have its origin on the datum surface, that is on the ellipsoid. In the classical (non-three-dimensional) approach the orthometric height H enters into horizontal networks only in the reduction of surface quantities to the geoid, therefore it is natural to take our local astronomic system at the initial point to have its origin on the geoid. Denoting quantities at the initial point by a zero subscript, we then see that the six datum position parameters are in this case the geodetic coordinates of the local astronomic origin (ϕ_o, λ_o, N^*_o) and the rotation angles required to define the transformation between the local geodetic and local astronomic systems (ξ_o, η_o, $\delta\alpha_o$).

2.3.2 Establishment of a Datum

We have seen that a datum is defined by assigning values either to the eight parameters (a, b, x_o, y_o, z_o, w_1, w_2, w_3) or to the eight parameters (a, b, ϕ_o, λ_o, N^*_o, ξ_o, η_o, $\delta\alpha_o$). However, an arbitrary set of values will not in general result in a satisfactory datum. We recall that it is important that a datum closely approximate the geoid over the area of the network for which it is a datum, and that the geocentric axes of the **geodetic** coordinate system be closely parallel to the Average Terrestrial axes, particularly that the axes of symmetry be parallel. The process of assigning values to the eight datum parameters in such a way that these characteristics are obtained is called <u>establishment of a datum</u>.

To begin with, in establishing a datum values are always assigned to the topocentric set (a, b, ϕ_o, λ_o, N^*_o, ξ_o, η_o, $\delta\alpha_o$) rather than the geocentric set (a, b, x_o, y_o, z_o, w_1, w_2, w_3) because it is the set which is related to the geodetic and astronomic measurements which we must use in establishing the datum. We see that we must somehow choose values for (a, b, ϕ_o, λ_o, N^*_o, ξ_o, η_o, $\delta\alpha_o$) so that the values of (N^*,ξ,η) elsewhere in the network are not excessive (the datum approximates the geoid), and so that $w_1 = w_2 = w_3 = 0$ (the axes are parallel). Additionally for networks of global extent we require that $x_o = y_o = z_o = 0$, in which case the datum is termed a geocentric datum. Otherwise the datum is a local datum.

The problem of approximating the geoid can be ignored, in which case the values

$$N^*_o = \xi_o = \eta_o = 0$$

are assigned, which forces the ellipsoid to intersect and be tangent to the geoid at the initial point.

The geoid can be approximated in two ways, by choosing values of (a, b, N^*_o, ξ_o, η_o) such that either values of (ξ, η) or values of N^* throughout the network are minimized (Vanicek, 1972). Note that values of (N^*, ξ, η) are available throughout the network only if some adjusted network already exists, which points up the iterative nature of datum establishment - a "best fitting" datum can be established only as an improvement on an already existing datum.

The classical method of "ensuring" that the axes of symmetry are parallel is to enforce the Laplace azimuth condition at the initial point, that is to assign a value to α_o according to

$$\delta\alpha_o = A_o - \alpha_o = \eta_o \tan \phi_o \qquad 2\text{-}44$$

where A_o is an observed astronomic azimuth. This condition forces the geodetic and astronomic meridians to be parallel at the initial point, and thus forces both axes of symmetry to lie in this common plane. However, the axes of symmetry can still be misaligned within the meridian plane. The solution to this dilemma has been to apply the Laplace condition at several geodetic meridians parallel to their corresponding astronomic meridians. In essence this constrains the adjusted network to compensate for misalignment of the datum, rather than ensuring that the datum minor axis is parallel to the earth's rotation axis. Note that enforcing the Laplace condition throughout the network presumes the existance of an adjusted network, which again points up the iterative nature of datum establishment.

2.3.3 The North American Datum

The iterative nature of datum establishment is illustrated by the history of the North American Datum.

Towards the close of the last century geodetic networks existed in several parts of North America, each defined on its own datum. The largest of these was the New England Datum established in 1879 with an initial point at Principio, Maryland. The New England Datum used the Clarke 1866 ellipsoid, still used by the North American Datum today.

By 1899 the U.S. Transcontinental Network linking the Atlantic and Pacific coasts was complete. When an attempt was made to join the newer networks to those of the New England Datum large discrepancies occurred. Therefore in 1901 the United States Standard Datum was established. The Clarke 1866 ellipsoid was retained from the New England Datum, but the initial point was moved from Principio to the approximate geographical centre of the U.S. at Meades Ranch, Kansas. The coordinates and azimuth at Meades Ranch were selected so as to cause minimum change in existing coordinates and publications (mainly in New England) while providing a better fit to the geoid for the rest of the continent.

Meanwhile additional networks were being established in the United States, Canada and Mexico. In 1913 Canada and Mexico agreed to accept Meades Ranch as the initial point for all North American networks, and the datum was renamed the North American Datum.

This eventually led to the readjustment, between 1927 and 1932, of all the North American networks then in existance. The 1901 coordinates of Meades Ranch and the Clarke 1866 ellipsoid remained unchanged, however the value of the geodetic azimuth was changed by about 5 arcseconds (Mitchell, 1948). Thus the new datum was called the 1927 North American Datum.

The definition of the North American Datum was not yet complete. It was only in 1948 that astronomic coordinates were observed at Meades Ranch, allowing specification of values for ξ_o, η_o . The final datum parameter was defined in 1967 when the U.S. Army Map Service chose a value of $N_o^* = 0$ at Meades Ranch for their astrogeodetic geoid [Fischer, et al 1967]. Table 2-1 lists the values assigned to the datum parameters for the North American Datum, and the date at which they were determined.

Since the 1927 readjustment many new networks have been added to what was then available. However, these new networks have been adjusted by "tacking them on" to previously adjusted networks, the latter being held fixed in the process. Until the recent advent of large fast digital computers it was impractical to consider readjusting all the networks on the continent again, consequently distortions have crept in to the networks, a notorious case being the 10 metres discrepancy which has been "drowned" in Lake Superior by international agreement. The day is fast approaching when a massive new readjustment and perhaps redefinition of the North American Datum will occur [Smith, 1971]. One landmark on this path is the International Symposium on Problems Related to the Redefinition of North American Geodetic Networks,

Table 2-1

PARAMETERS DEFINING THE

1927 NORTH AMERICAN DATUM

Parameter	Symbol	Value	Date Adopted
Clarke 1866 Ellipsoid semi-major axis	a =	6378206.4 metres	1879
Clarke 1866 Ellipsoid semi-minor axis	b =	6356583.8 metres	
Initial Point Latitude of Meade's Ranch	ϕ_o =	39° 13' 26".686 N	1901
Initial Point Longitude of Meade's Ranch	λ_o =	98° 32' 30".506 W	
Initial Point Azimuth (to Waldo)	α_{oi} =	75° 28' 9".64 (clockwise from south)	1927
Initial Point Meridian Deflection Component	ξ_o =	-1.02"	1948
Initial Point Prime Vertical Deflection Component	η_o =	-1.79"	
Initial Point Geoid Height	N^*_o =	0	1967

Table 2-2

TRANSLATION COMPONENTS

	x_o	y_o	z_o	σ_{xo}	σ_{yo}	σ_{zo}
Merry & Vanicek	-28.7	150.5	179.9	1.7	1.0	1.2
Krakiwsky et al.	-35	164	186	2	3	3

May 1974 at the University of New Brunswick.

The North American Datum is a local datum, that is its geometrical centre does not coincide with the origin of the Average Terrestrial system. Because of the distortions in the networks just mentioned, determinations of x_o, y_o, z_o vary depending on the locations at which they are measured. Two recent sets of values obtained by different methods are listed in Table 2-2. Merry and Vanicek [1973] used data within 1000 km of Meades Ranch. Krakiwsky et al [1973] used data from New Brunswick and Nova Scotia. The discrepancies of order 10 metres likely reflect the distortions which exist in the present North American networks.

2.3.4 Datum Transformations

If the curvilinear coordinates of an observing station referring to one particular datum are given, then a problem which often occurs is to obtain the curvilinear coordinates for the station referred to another datum.

In transforming coordinates from one datum to another it is necessary to account for two items:

a) the location of the geometric centres of each reference ellipsoid with respect to the centre of gravity of the earth, or with respect to each other,

b) the difference in size and shape between the ellipsoids.

It is usually assumed that the axes of both datums are parallel to the axes of the average terrestrial system.

Consider the ellipsoids with sizes and shapes defined by (a_1, b_1) and (a_2, b_2) (or alternatively (a_1, f_1) and (a_1, f_2), where $f=(a-b)/a$.) and with locations of the geometric centres with respect to the centre of gravity defined by

$$(\bar{r}_o)_1 = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}_1$$

and

$$(\bar{r}_o)_2 = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}_2 .$$

Let us define the coordinates of a point referred to the first ellipsoid as (ϕ_1, λ_1, h_1). We want to find the coordinates of the same point, referred to the second ellipsoid (ϕ_2, λ_2, h_2).

The average terrestrial coordinates of the point are given by equation 2-32:

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}_1 + \begin{bmatrix} (N_1 + h_1) \cos\phi_1 \cos\lambda_1 \\ (N_1 + h_1) \cos\phi_1 \sin\lambda_1 \\ (N_1 b_1^2/a_1^2 + h_1) \sin\phi_1 \end{bmatrix} .} \quad 2\text{-}45$$

But the average terrestrial coordinates are not affected by a datum transformation, therefore

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}_2 + \begin{bmatrix} (N_2 + h_2) \cos\phi_2 \cos\lambda_2 \\ (N_2 + h_2) \cos\phi_2 \sin\lambda_2 \\ (N_2 b_2^2/a_2^2 + h_2) \sin\phi_2 \end{bmatrix} .} \quad 2\text{-}46$$

There are two methods for obtaining (ϕ_2, λ_2, h_2). The first method, called the <u>iterative</u> method is to find the average terrestrial coordinates directly from equation 2-45, and then to invert equation 2-46 to find (ϕ_2, λ_2, h_2), using the iterative method described in section 2.2.6.

The second method, called the <u>differential</u> method can be applied when the parameter differences (δa, δf, δx_o, δy_o, δz_o) between the two datums is small enough that we can use the Taylor's series linear approximation. Taking the total differential of equation 2032, keeping the average terrestrial coordinates invariant, and setting the differential quantities equal to differences between the datums we have

$$\begin{bmatrix} \delta x_o \\ \delta y_o \\ \delta x_o \end{bmatrix} + J \begin{bmatrix} \delta\phi \\ \delta\lambda \\ \delta h \end{bmatrix} + B \begin{bmatrix} \delta a \\ \delta f \end{bmatrix} = 0 \qquad 2\text{-}47$$

where

$$J = \begin{bmatrix} -(M+h)\sin\phi\cos\lambda, & -(N+h)\cos\phi\sin\lambda, & \cos\phi\cos\lambda \\ -(M+h)\sin\phi\sin\lambda, & (N+h)\cos\phi\cos\lambda, & \cos\phi\sin\lambda \\ (M+h)\cos\phi, & 0, & \sin\phi \end{bmatrix} \qquad 2\text{-}48$$

$$B = \begin{bmatrix} N\cos\phi\cos\lambda/a, & M\sin^2\phi\cos\phi\cos\lambda/(1-f) \\ N\cos\phi\sin\lambda/a, & M\sin^2\phi\cos\phi\sin\lambda/(1-f) \\ N(1-f)^2\sin\phi/a, & (M\sin^2\phi - 2N)\sin\phi(1-f) \end{bmatrix} \qquad 2\text{-}49$$

$$M = a(1-f)^2/(\cos^2\phi + (1-f)^2\sin^2\phi)^{3/2} \qquad 2\text{-}50$$

Solving for the coordinate differences

$$\begin{bmatrix} \delta\phi \\ \delta\lambda \\ \delta h \end{bmatrix} = -J^{-1}\left\{ \begin{bmatrix} \delta x_o \\ \delta y_o \\ \delta z_o \end{bmatrix} + B \begin{bmatrix} \delta a \\ \delta f \end{bmatrix} \right\} \qquad 2\text{-}51$$

where

$$J^{-1} = \begin{bmatrix} -\sin\phi\cos\lambda/(M+h), & -\sin\phi\sin\lambda/(M+h), & \cos\phi/(M+h) \\ -\sin\lambda/(N+h)\cos\phi, & \cos\lambda/(N+h)\cos\phi, & 0 \\ \cos\phi\cos\lambda, & \cos\phi\sin\lambda, & \sin\phi \end{bmatrix} \qquad 2\text{-}52$$

Note that the matrices can be evaluated in either of the two coordinate systems, since the differences in quantities has been assumed small. Further it is reasonable to simplify the evaluation of the matrices by using the spherical approximation ($f = 0$, $N = M = N+h = M+h = a$) in which case we obtain the transformation equations of Heiskanen and Moritz (1967, equation 5-55).

Table 2-3 shows an example of datum transformation computations. This particular example transforms the coordinates of a station in Dartmouth, Nova Scotia from the 1927 North American Datum ("Old"Datum) to the 1950 European Datum ("New" Datum). The datum translation components used are those given by Lambeck [1971]. Both the iterative method of equations 2-45 and 2-46, and the differential method of equation 2-51 were used. The discrepancies between the two results are about 0.4 meters in latitude, 0.3 meters in longitude, and 0.2 meters in height.

2.4 TERRESTRIAL TOPOCENTRIC SYSTEMS

In the introduction it was stated that terrestrial topocentric systems are defined as follows:

a) the origin is at a point near the surface of the earth,

b) the primary plane is the plane tangential to the earth's surface at the point,

c) the primary axis is the north point,

d) the systems are left-handed.

The last two specifications present no problems. However, "the surface of the earth" can be interpreted in three ways to mean the earth's physical surface, the earth's equipotential surface, or the

Table 2-3.

EXAMPLE OF DATUM TRANSFORMATIONS

Parameter		"Old" Datum	"New" Datum	"New"-"Old"
Given:				
semi-major axis	a	6378206.4 meters	6378388.0	181.6
flattening	f	1/294.98	1/297.0	-2.3057×10^{-5}
offset from geocentre (from Lambeck [1971])	x_o	-25.8	-64.5	-38.7
	y_o	168.1	-154.8	-322.9
	z_o	167.3	-46.2	-213.5
observer's coordinates	ϕ_1	44.683°N	?	
	λ_1	63.612°W	?	
	h_1	37.46 meters	?	
Solution by Iterative Method:				
observer's coordinates in average Terrestrial System (Equation2-45)	x	2018917.91	2018917.91	
	y	-4069107.35	-4069107.35	
	z	4462360.64	4462360.64	
observer's coordinates (Equation 2-46)	ϕ_2		44.684770°N	
	λ_2		63.609752°W	
	h_2		-259.73 meters	
Solution by differential Method:				
change in semi-major axis	δa			181.6
change in flattening	δf			-2.3057×10^{-5}
change in offsets from geocenter	δx_o			-38.7
	δy_o			-322.9
	δz_o			-213.5
observer's coordinates (Equation 2-50).	ϕ_2		44.684766°N	
	λ_2		63.609749°W	
	h_2		-259.92 meters	

surface of a reference ellipsoid. It is not practical to define a coordinate system in terms of a plane tangential to the earth's physical surface. Two kinds of terrestrial topocentric coordinate systems can be defined, however; The system in which the primary pole is the normal to the equipotential surface at the observation station is called a local astronomic system; The system in which the primary pole is the ellipsoid normal passing through the observation station is called a local geodetic system [Krakiwsky 1968].

2.4.1 Local Astronomic System

A local astronomic (L.A.) system is specified:

a) The origin is at the observation station.

b) The primary pole (z-axis) is the normal to the equipotential surface (the gravity vertical) at the observation station. The primary plane is the plane containing the origin and perpendicular to the gravity vertical.

c) The primary axis (x-axis) is the intersection of the primary plane and the plane containing the average terrestrial pole and the observation station, and is called the astronomic north.

d) The y-axis is directed east to form a left-handed system.

The position vector of an observed station l, expressed in the local astronomic system of the observation station k, is given by

$$\boxed{(\bar{r}_{kl})_{L.A.} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{L.A.} = r_{kl} \begin{bmatrix} \cos v_{kl} \cos A_{kl} \\ \cos v_{kl} \sin A_{kl} \\ \sin v_{kl} \end{bmatrix},} \qquad 2\text{-}54$$

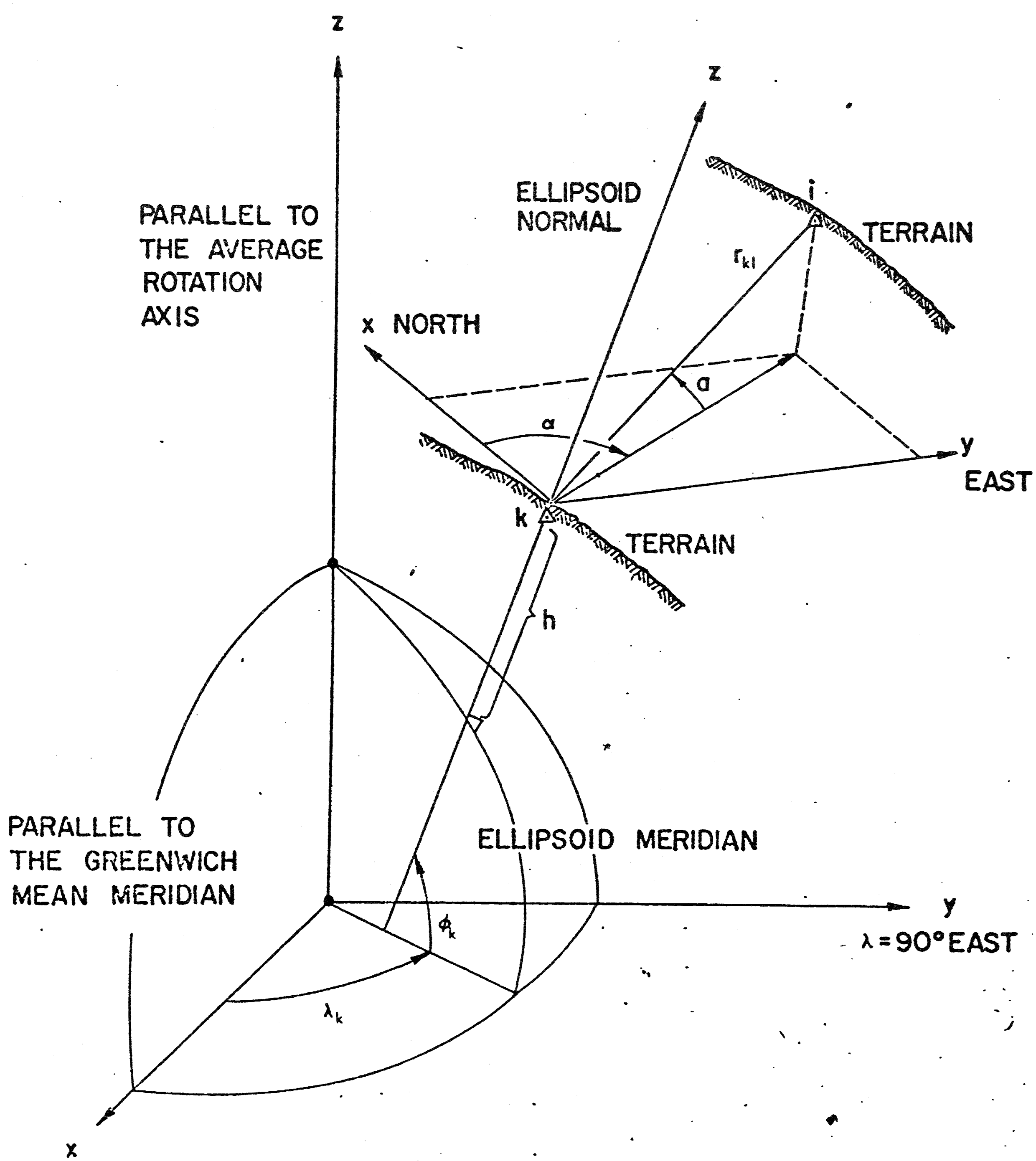


Figure 2-12

GEODETIC AND LOCAL GEODETIC COORDINATE SYSTEMS where r_{kl} is the terrestrial spatial distance, v_{kl} the vertical angle, and A_{kl} the astronomic azimuth.

Note that the relationship of the local astronomic system to the average terrestrial system is given by the astronomic latitude Φ_k and longitude Λ_k only after the observed quantities Φ_k, Λ_k, A_{kl} have been corrected for polar motion. Thus the position vector $\bar{r}_{kl}$ of 2-54 expressed in the average terrestrial system is:

$$\boxed{\left(\bar{r}_{kl}\right)_{A.T.} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = R_3(180^\circ-\Lambda_k)\ R_2(90^\circ-\Phi_k)P_2 \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{L.A.}} \qquad 2\text{-}55$$

where the reflection matrix

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad 2\text{-}56$$

accomplishes the transformation from a left-handed system into a right-handed system, while the rotation matrices

$$R_2 = \begin{bmatrix} \cos\ (90^\circ-\Phi_k) & 0 & -\sin\ (90^\circ-\Phi_k) \\ 0 & 1 & 0 \\ \sin\ (90^\circ-\Phi_k) & 0 & \cos\ (90^\circ-\Phi_k) \end{bmatrix} \qquad 2\text{-}57$$

and

$$R_3 = \begin{bmatrix} \cos(180^\circ-\Lambda_k) & \sin(180^\circ-\Lambda_k) & 0 \\ -\sin(180^\circ-\Lambda_k) & \cos(180^\circ-\Lambda_k) & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad 2\text{-}58$$

bring the three axes of the local astronomic system parallel to the corresponding axes in the average terrestrial system.

The inverse transformation is

$$(\bar{r}_{kl})_{L.A.} = [R_3\ (180^\circ-\Lambda_k)\ R_2(90^\circ-\Phi_k)\ P_2]^{-1}\ (\bar{r}_{kl})_{A.T.} \qquad 2\text{-}59$$

$$= P_2 R_2 (\Phi_k - 90^\circ) \, R_3(\Lambda_k - 180^\circ) \, (\bar{r}_{kl})_{A.T.} \quad . \qquad 2\text{-}60$$

Note that so far no translations have taken place. We have merely rotated the position vector $(\bar{r}_{kl})$ of station l with respect to station k into the average terrestrial system. If the position vector of station k with respect to the centre of gravity in the average terrestrial system is $(\bar{R}_k)_{A.T.}$, then the total position vector $\bar{R}_l$ of the observed station l with respect to the centre of gravity in the average terrestrial system is given by

$$(\bar{R}_l)_{A.T.} = (\bar{R}_k)_{A.T.} + (\bar{r}_{kl})_{A.T.} \quad . \qquad 2\text{-}61$$

The unit vectors $\hat{u}_x$, $\hat{u}_y$, $\hat{u}_z$ directed along the axes of the local astronomic system have the following components in the average terrestrial system:

$$\hat{u}_x = R_3 (180^\circ - \Lambda) \, R_2 (90^\circ - \Phi) \, P_2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\boxed{\hat{u}_x = \begin{bmatrix} -\sin\Phi \ \cos\Lambda \\ -\sin\Phi \ \sin\Lambda \\ \cos\Phi \end{bmatrix}} \ , \qquad 2\text{-}62$$

$$\hat{u}_y = R_3 (180^\circ - \Lambda) \, R_2 (90^\circ - \Phi) \, P_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\boxed{\hat{u}_y = \begin{bmatrix} -\sin \Lambda \\ \cos \Lambda \\ 0 \end{bmatrix}} \ , \qquad 2\text{-}63$$

$$\hat{u}_z = R_3\ (180° - \Lambda)\ R_2\ (90° - \Phi)\ P_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\boxed{\hat{u}_z = \begin{bmatrix} \cos\Phi \quad \cos\Lambda \\ \cos\Phi \quad \sin\Lambda \\ \sin\Phi \end{bmatrix} .} \qquad \text{2-64}$$

The local astronomic coördinate system is unique for every observation point. Because of this fact, this system is the basis for treating terrestrial three-dimensional measurements at several stations together in one solution.

2.4.2. Local Geodetic System

A local geodetic (L.G.) system is specified (see Figure 2-12):

a) The origin lies along the ellipsoidal normal passing through the observation station. Note that in principle the origin may lie anywhere along the ellipsoidal normal. In practice it is chosen to be at the observation station, at the ellipsoid, or at the intersection of the ellipsoidal normal with the geoid.

b) The primary pole (z-axis) is the ellipsoidal normal. The primary plane is the plane containing the origin and perpendicular to the primary pole.

c) The primary axis (x-axis) is the intersection of the primary plane and the plane containing the semi-minor axis of the ellipsoid and the origin, and is called the geodetic north.

d) The y-axis is directed east to form a left-handed system.

Transformations between local geodetic and local astronomic systems sharing a common origin can be expressed in terms of the angle between the ellipsoid normal and gravity vertical (the deflection of the vertical) and the angle between the geodetic north and astronomic north. Given the meridian and prime vertical deflection components ξ, η respectively, and the geodetic and astronomic azimuths α, A to a particular point, then a vector in the local astronomic system is transformed into a vector in the local geodetic system by

$$(\bar{r}_{kl})_{L.G.} = R_3\,(A-\alpha)\;R_2(-\xi)\;R_1(\eta)\;(\bar{r}_{kl})_{L.A.} \qquad 2.65$$

Note that the order in which the rotations are performed in this case is not important, since the angles ξ, η $(A - \alpha)$ are small enough that their rotation matrices can be assumed to commute. Note also that if the Laplace condition is enforced at the origin of these local systems, we have

$$A - \alpha = (\Lambda - \lambda)\sin\phi = \eta\tan\phi$$

If the origin is not at the observation station, the position vector $\bar{R}_k$ in 2-61 would refer to the origin, not the observation station. That is, for a point on the geoid, the computation of (x_k, y_k, z_k) is made from (ϕ_k, λ_k, N_k) (geoid undulation), while on the ellipsoid $(\phi_k, \lambda_k, 0)$ are used. Note that when a small region of the earth is taken as a plane it is a local geodetic system that is implied.

Similar to equations 2-54 and 2-55 the position vector from observing station k to observed station l is given by

$$\boxed{(\bar{r}_{kl})_{L.G.} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{L.G.} = r_{kl} \begin{bmatrix} \cos a_{kl} & \cos\alpha_{kl} \\ \cos a_{kl} & \sin\alpha_{kl} \\ \sin a_{kl} & \end{bmatrix}} \qquad 2\text{-}66$$

and

$$(\bar{r}_{kl})_{G.} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_G = R_3(180^\circ-\lambda_k)\ R_2(90^\circ-\phi_k)\ P_2 \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{L.G.} \qquad \text{2-67}$$

where (a, α, r) are the geodetic altitude, azimuth and range, and (φ, λ) are the geodetic latitude and longitude. Note that the geodetic system (G) and the average terrestrial system (A.T.) are related by equation 2-4

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} + \begin{bmatrix} x \\ y \\ z \end{bmatrix}_G ,$$

where (x_o, y_o, z_o) are the translation components of the origin of the geodetic system in the average terrestrial system.

The unit vectors corresponding to the three Cartesian axes in the local geodetic system are

$$\hat{u}_x = \begin{bmatrix} -\sin\phi \ \cos\lambda \\ -\sin\phi \ \sin\lambda \\ \cos\phi \end{bmatrix} , \qquad \text{2-68a}$$

$$\hat{u}_y = \begin{bmatrix} -\sin\lambda \\ \cos\lambda \\ 0 \end{bmatrix} , \qquad \text{2-68b}$$

$$\hat{u}_z = \begin{bmatrix} \cos\phi \ \cos\lambda \\ \cos\phi \ \sin\lambda \\ \sin\phi \end{bmatrix} . \qquad \text{2-68c}$$

2.5 SUMMARY OF TERRESTRIAL SYSTEMS

In this chapter we have precisely defined five specific terrestrial coordinate systems:

a) Average Terrestrial (A.T.),

b) Instantaneous Terrestrial (I.T.),

c) Geodetic (G),

d) Local Astronomic (L.A.),

e) Local Geodetic (L.G.),

of which the first three are geocentric and the last two topocentric. Table 2-4 summarizes the planes, poles and axes defining these systems. We have also precisely defined four kinds of coordinates:

a) Cartesian (x, y, z) - used by all systems,

b) Curvilinear (ϕ, λ, h) - used by Geodetic system,

c) Curvilinear (v, A, r) - used by Local Astronomic system,

d) Curvilinear (a, α, r) - used by Local Geodetic system.

Finally we have defined the principal transformations between these coordinates and coordinate systems. Figure 2-13 lists the equation numbers which define these transformations, which are tabulated in Table 2-5.

Table 2-4.

REFERENCE POLES, PLANES AND AXES DEFINING TERRESTRIAL COORDINATE SYSTEMS

System	Reference Poles		Reference Planes		Handedness
	Primary (z-axis)	Secondary (x-axis)	Primary ($\perp$ to Primary Pole)	Secondary	
Average Terrestrial	Average Terrestrial Pole (CIO)	Intersection of Primary and Secondary Planes	Average Terrestrial equator containing centre of gravity.	Greenwich mean meridian	right
Instantaneous Terrestrial	Instantaneous Terrestrial Pole		Instantaneous Terrestrial equator.	Greenwich mean meridian	right
Geodetic	Semi-minor axis (parallel to terrestrial pole)		Parallel To Average Terrestrial equator	Parallel to Greenwich mean meridian	right
Local Astronomic	Gravity Vertical at Station		Local Horizon	Astronomic Meridian of station.	left
Local Geodetic	Ellipsoidal Normal at Station.		Tangent Plane	Coincident with Geodetic Meridian of station.	left

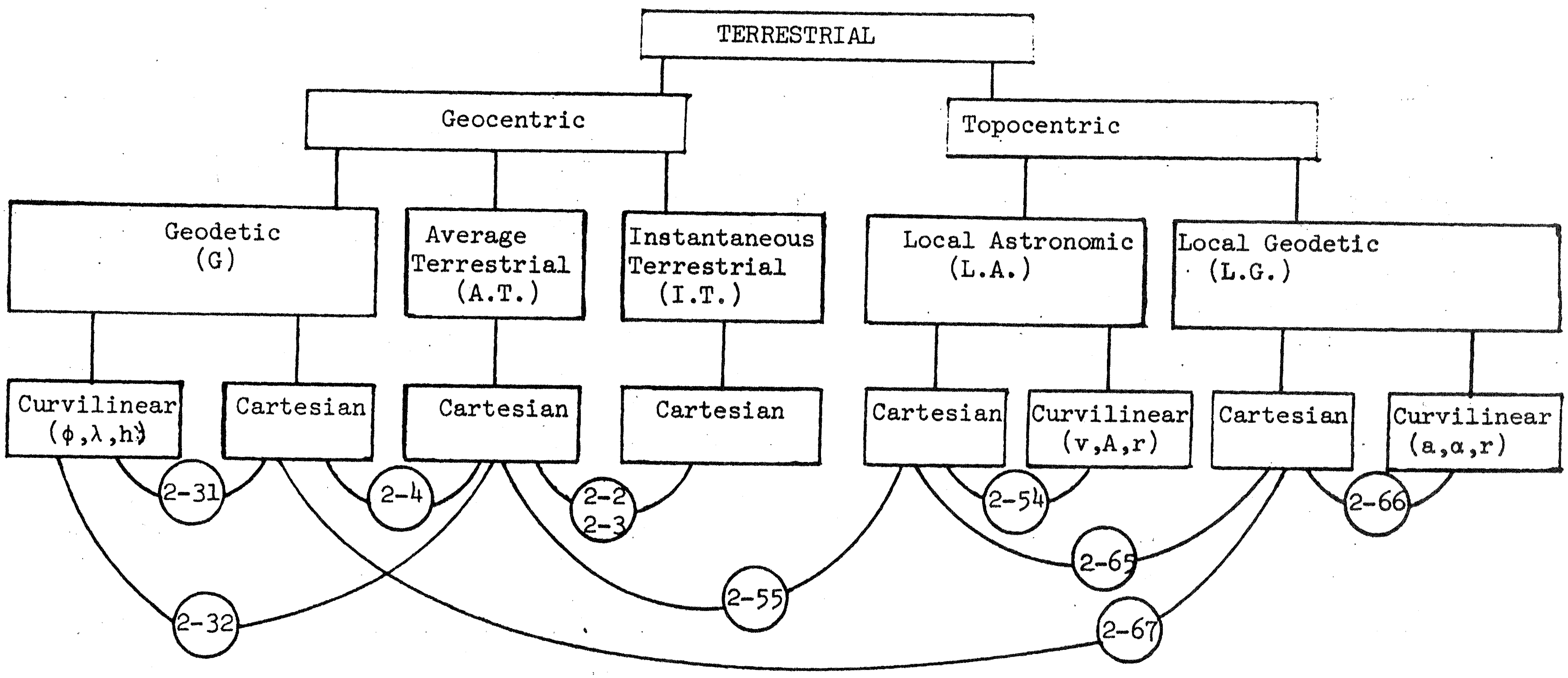


Figure 2-13.

EQUATIONS RELATING TERRESTRIAL SYSTEMS

Table 2-5.

TRANSFORMATIONS AMONG TERRESTRIAL COORDINATE SYSTEMS.

Final System ↓ / Original System →	Average Terrestrial	Instantaneous Terrestrial	Geodetic	Local Astronomic	Local Geodetic
Average Terrestrial	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.}$	$R_2(-x_p)R_1(-y_p)$	$+\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}$	$R_3(180°-\Lambda)R_2(90°-\Phi)P_2$	via Geodetic
Instantaneous Terrestrial	$R_1(y_p)R_2(x_p)$	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{I.T.}$	via Average Terrestrial	via Average Terrestrial	via Geodetic
Geodetic	$-\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}$	via Average Terrestrial	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{G}$	via Local Geodetic	$R_3(180°-\lambda)R_2(90°-\phi)P_2$
Local Astronomic	$P_2R_2(\Phi-90°)R_3(\Lambda-180°)$	via Average Terrestrial	via Local Geodetic	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{L.A.}$	$R_2(+\xi)\ R_1(+\eta)$
Local Geodetic	via Geodetic	via Average Terrestrial	$P_2R_2(\phi-90°)R_3(\lambda-180°)$	$R_1(-\eta)R_2(-\xi)$	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{L.G.}$

Note: Φ, Λ have been corrected for polar motion.

3. CELESTIAL COORDINATE SYSTEMS

Celestial coordinate systems are used to define the coordinates of celestial bodies such as stars. The distance from the earth to the nearest star is more than 10^9 earth radii, therefore, the dimensions of the earth (indeed of the solar system) are almost negligible compared to the distances to the stars. A second consequence of these great distances is that, although the stars themselves are believed to be moving at velocities near the velocity of light, to an observer on the earth this motion is perceived to be very small, very rarely exceeding one arcsecond per year. Therefore, the relationship between the earth and stars can be closely approximated by considering the stars all to be equidistant from the earth, on the surface of the celestial sphere, the dimension of which is so large that the earth (and indeed the solar system) can be considered as a dimensionless point at the centre. Although this point may be dimensionless, relationships between directions on the earth and in the solar system can be extended to the celestial sphere.

The earth's rotation axis is extended outward to intersect the celestial sphere at the north celestial pole (NCP) and south celestial pole (SCP). The earth's equatorial plane extended outward intersects the celestial sphere at the celestial equator. The gravity vertical at a station on the earth is extended upwards to intersect the celestial sphere at the zenith (Z), and downwards to intersect at the nadir (N). The plane of the earth's orbit around the sun (the ecliptic plane) is extended outward to intersect the celestial sphere at the ecliptic. The line of intersection between the earth's equatorial plane and the ecliptic plane is extended outwards to intersect the celestial sphere at the vernal equinox or first point of Aries, and the autumnal equinox. The vernal equinox is denoted by the symbol ♈, and is the point at which the sun crosses the celestial equator from south to north.

There are two fundamental differences between celestial systems and terrestrial or orbital systems. First, only directions and not distances are considered in celestial coordinate systems. In effect this means that the celestial sphere can be considered the unit sphere, and all vectors dealt with are unit vectors. The second difference is related to the first, in that the celestial geometry is spherical rather than ellipsoidal as in terrestrial and orbital systems, which simplifies the mathematical relationships involved.

As discussed in the introduction, there are four main celestial coordinate systems, called the ecliptic, right ascension, hour angle, and horizon. Sometimes the right ascension and hour angle systems are referred to collectively as equatorial systems. We will begin this chapter by discussing each of these systems in turn.

We noted above that the celestial sphere is only an approximation of the true relationship between the stars and an observer on the earth.

Therefore, like all approximations, there are a number of corrections which must be made to precisely represent the true relationship. These corrections represent the facts that the stars are not stationary points on the celestial sphere but are really moving (proper motion); the earth's rotation axis is not stationary with respect to the stars (precession and nutation); the earth is displaced from the centre of the celestial sphere, which is at the sun (parallax); the earth is in motion around the centre of the celestial sphere (aberration); and directions measured through the earth's atmosphere are bent by refraction. All these effects will be discussed in section 3.5 in terms of variations in the right ascension system.

3.1 THE ECLIPTIC SYSTEM

The ecliptic (E) system is specified as follows (see Figure 3-1):

a) The origin is heliocentric (at the centre of the sun).

b) The primary plane is the ecliptic plane (the plane of the earth's orbit) and the primary pole (z-axis) is the north ecliptic pole (NEP).

c) The primary axis (x-axis) is the vernal equinox.

d) The y-axis is chosen to make the system right-handed.

The ecliptic system is the celestial system which is closest to being inertial, that is motionless with respect to the stars. However, due to the effect of the planets on the sun-earth system, the ecliptic plane is slowly rotating (at $0''.5$ per year) about a slowly moving axis of rotation.

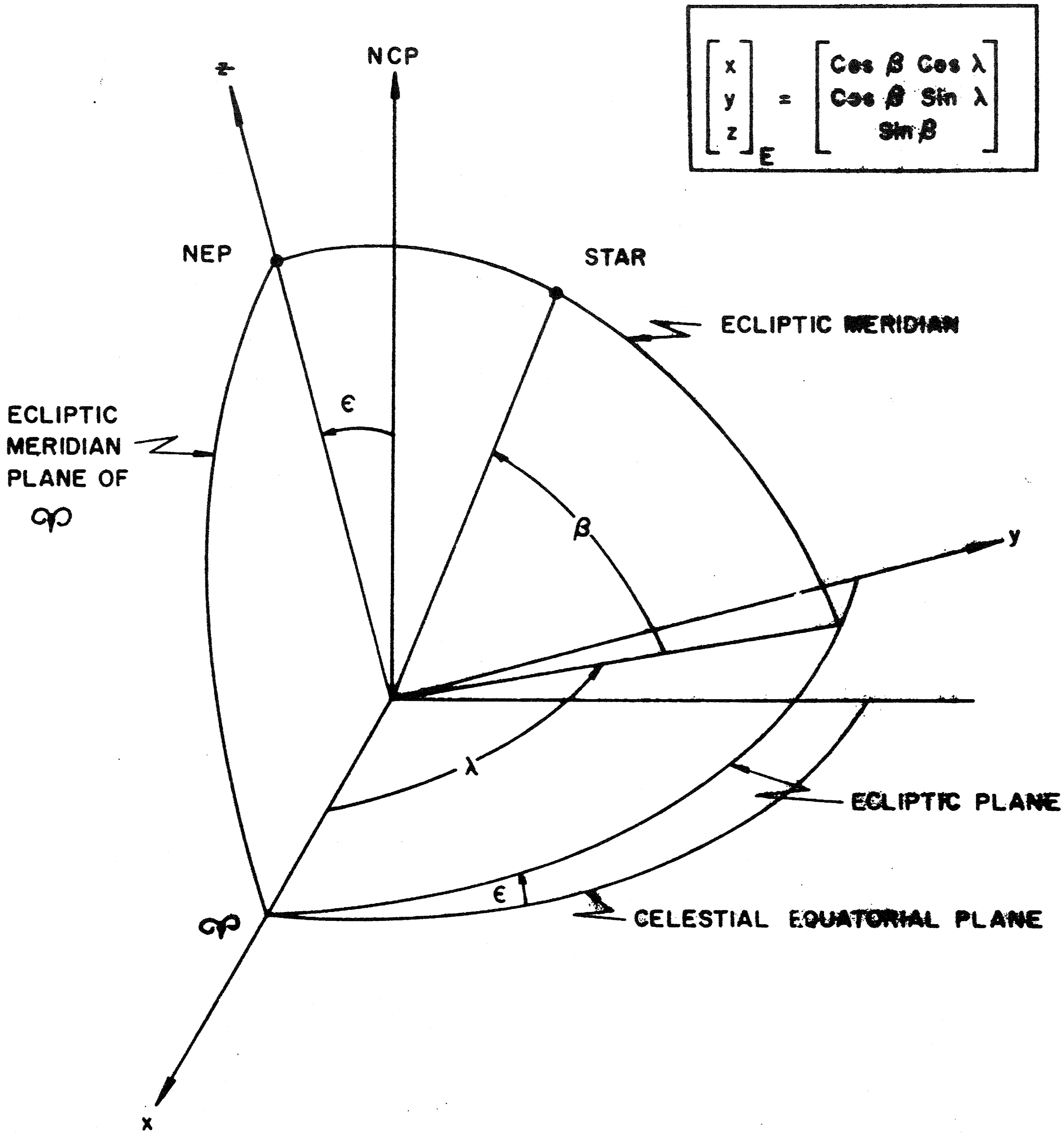


Figure 3-1. ECLIPTIC SYSTEM

The ecliptic meridian is the great circle which contains the ecliptic poles and the celestial body in question while the ecliptic meridian of ♈ contains the vernal equinox. Ecliptic latitude β is the angle from the ecliptic and in the ecliptic meridian to the line connecting the origin to the body. Ecliptic longitude λ is the angle of the ecliptic meridian of the body measured eastwards in the ecliptic plane from the vernal equinox. The unit vector to a celestial body in the celestial system is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_E = \begin{bmatrix} \cos\beta \ \cos\lambda \\ \cos\beta \ \sin\lambda \\ \sin\beta \end{bmatrix} ,} \qquad 3\text{-}1$$

and the angles are related to the Cartesian components by

$$\beta = \sin^{-1} z \ , \qquad 3\text{-}2$$

$$\lambda = \tan^{-1} \frac{y}{x} \ . \qquad 3\text{-}3$$

3.2 THE RIGHT ASCENSION SYSTEM

The right ascension (RA) system is specified as follows (see Figure 3-2).

a) The origin is heliocentric.

b) The primary plane is the equatorial plane, and the primary pole (z-axis) the north celestial pole (NCP).

c) The primary axis (x-axis) is the vernal equinox.

d) The y-axis is chosen to make the system right-handed.

The right ascension system is the most important celestial system. It

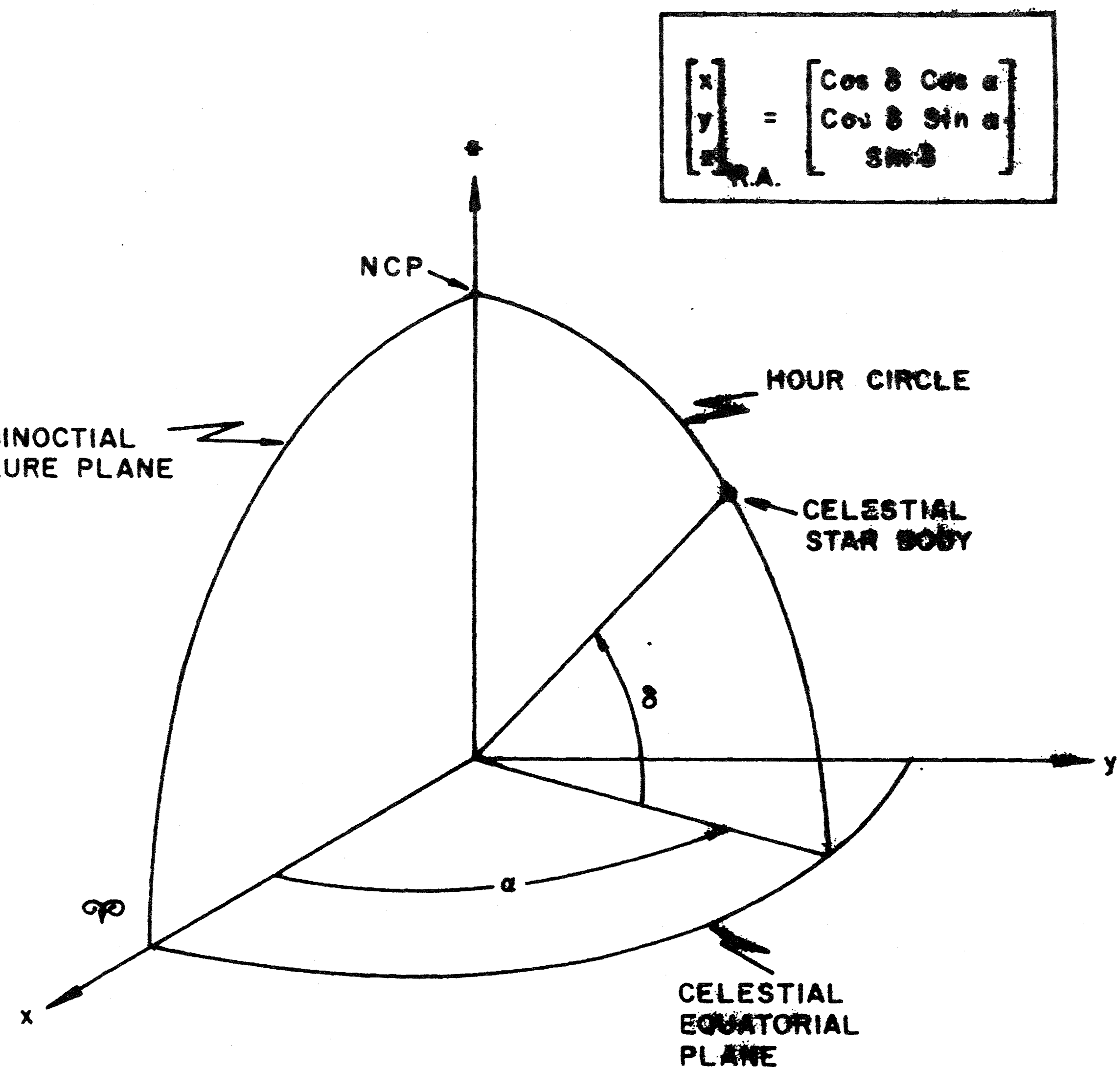


Figure 3-2. **RIGHT ASCENSION SYSTEM**

is in this system that star and satellite coordinates are published, and it serves as the connection between terrestrial, celestial and orbital systems.

The secondary plane contains the north celestial pole and the vernal equinox and is called the <u>equinoctial colure plane</u>. The <u>hour circle</u> is the great circle containing the celestial poles and the body in question. The <u>declination</u> δ of a body is then the angle between the celestial equator and a line joining the origin to the body. The <u>right ascension</u> α is the angle measured in the equatorial plane eastwards from the vernal equinox to the hour circle passing through the body in question. The unit vector describing the direction of a body in the right-ascension system is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{R.A.} = \begin{bmatrix} \cos\delta \ \cos\alpha \\ \cos\delta \ \sin\alpha \\ \sin\delta \end{bmatrix}.} \qquad 3\text{-}4$$

The right ascension system is related to the ecliptic system by the acute angle between the ecliptic and celestial equator, called the <u>obliquity of the ecliptic</u>, and denoted ε. Therefore

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{R.A.} = R_1(-\varepsilon) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{E}.} \qquad 3\text{-}5$$

3.3 THE HOUR ANGLE SYSTEM

The hour angle (HA) system is specified as follows (see Figure 3-3):

a) The origin is heliocentric.

b) The primary plane is the equatorial plane.

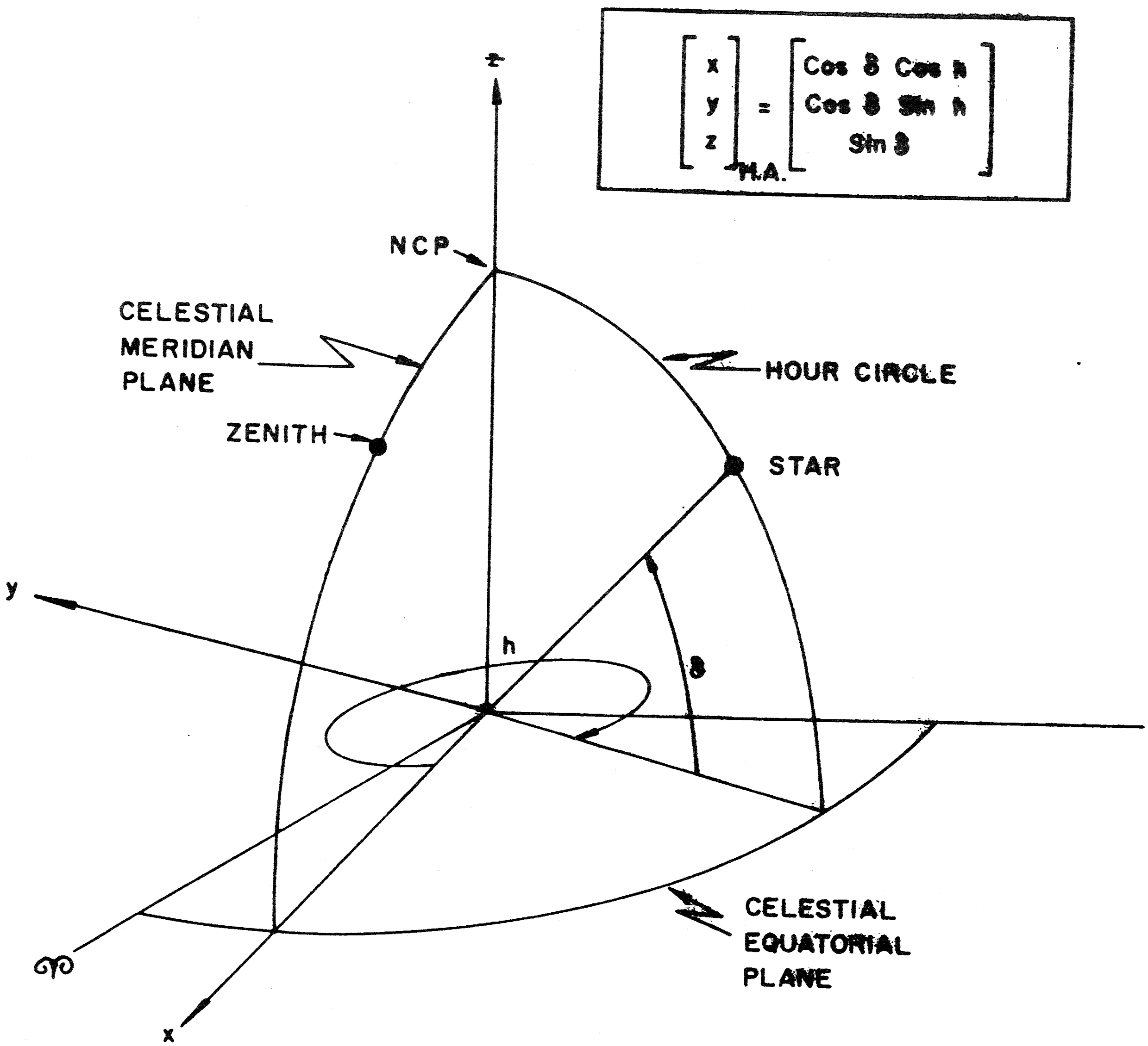


Figure 3-3. **HOUR ANGLE SYSTEM**

c) The secondary plane is the celestial meridian plane of the observer. The primary axis (x-axis) is the intersection between the equatorial and observer's celestial meridian planes.

d) The y-axis is chosen so that the system is left-handed.

The hour angle system rotates with the observer.

The <u>hour angle</u> h is the angle measured westward in the equatorial plane from the observer's celestial meridian to the hour circle of the body in question. The angle measured up from the equatorial plane to the line directed from the origin towards the body is the <u>declination</u> δ.

The unit vector describing the direction of a celestial body in the hour angle system is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{H.A.} = \begin{bmatrix} \cos\delta \cos h \\ \cos\delta \sin h \\ \sin\delta \end{bmatrix} .} \qquad 3\text{-}6$$

We have so far defined four meridians on the celestial sphere: that containing the vernal equinox (the equinoctial colure); the Greenwich meridian; that containing the observer (the celestial meridian); and that containing the star (the hour circle). Figure 3-4 shows the relationships between these meridians.

From the vernal equinox counterclockwise to

a) the Greenwich meridian is called Greenwich Sidereal Time (GST),

b) the celestial meridian is called Local Sidereal Time (LST),

c) the hour circle is called the right ascension (α).

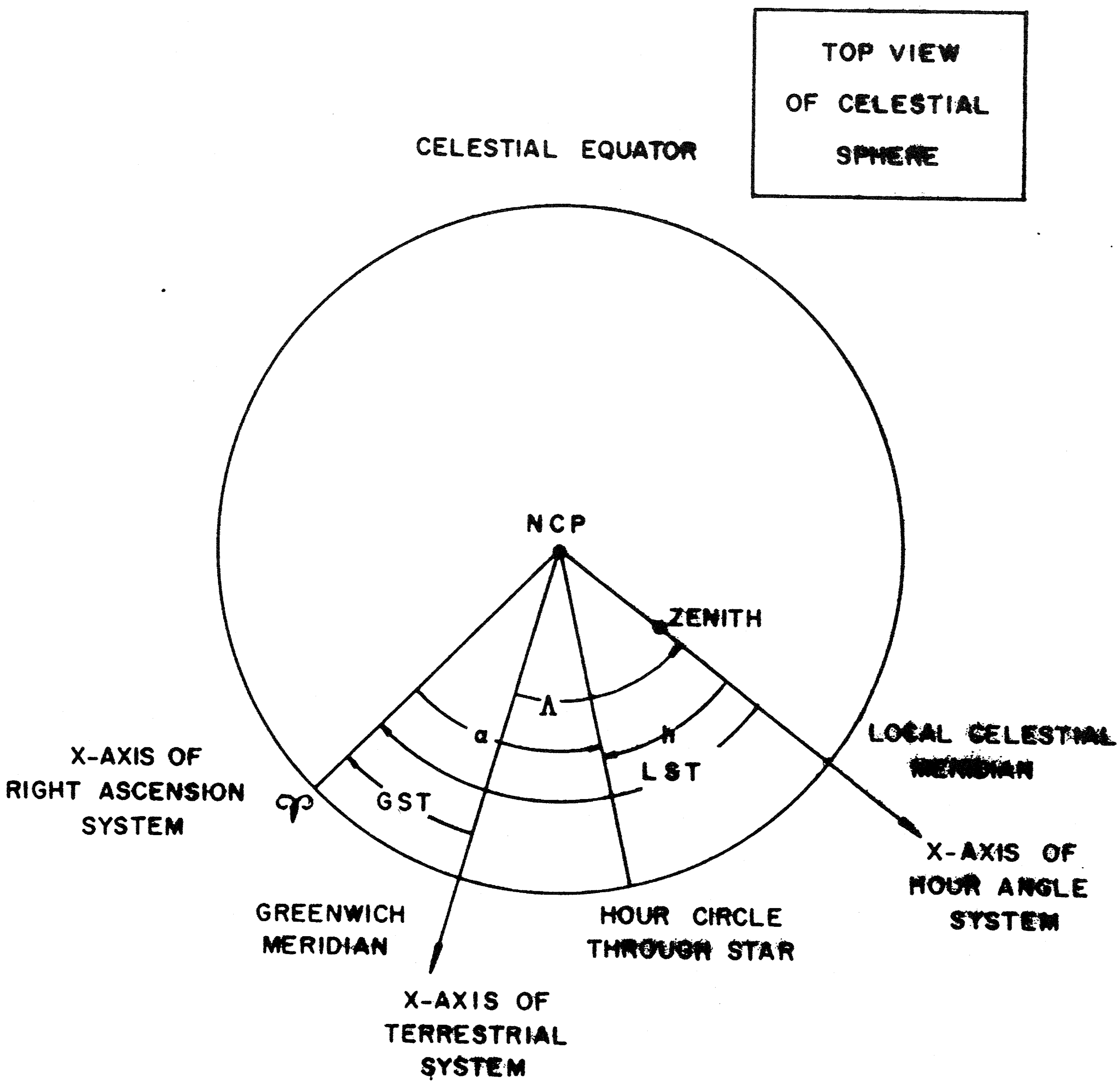


Figure 3-4.

TIME, LONGITUDE, AND RIGHT ASCENSION

From the Greenwich meridian counterclockwise to the celestial meridian is called the astronomic longitude (Λ). From the celestial meridian clockwise to the hour circle is called the hour angle (h). Therefore,

$$LST = GST + \Lambda \qquad 3\text{-}7$$

$$LST = h + \alpha \qquad 3\text{-}8$$

and

$$h = GST + \Lambda - \alpha \quad . \qquad 3\text{-}9$$

The hour angle system is related to the right ascension system by the local sidereal time (LST). That is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{HA} = P_2 \, R_3 \, (LST) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{RA} \quad .} \qquad 3\text{-}10$$

3.4 THE HORIZON SYSTEM

The horizon (H) system is specified as follows (see Figure 3-5).

a) The origin is heliocentric.

b) The primary pole (z-axis) is the observer's zenith (gravity vertical). The primary plane is the observer's horizon.

c) The primary axis (x-axis) is the north point.

d) The y-axis is chosen so that the system is left-handed.

The <u>horizon system</u> is used to describe the position of a celestial body in a system peculiar to a topocentrically located observer, similar to the local astronomic system described in the chapter on terrestrial systems. The main difference is that the origin of the horizon system is heliocentric instead of topocentric.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_H = \begin{bmatrix} \cos a \;\; \cos A \\ \cos a \;\; \sin A \\ \sin a \end{bmatrix}$$

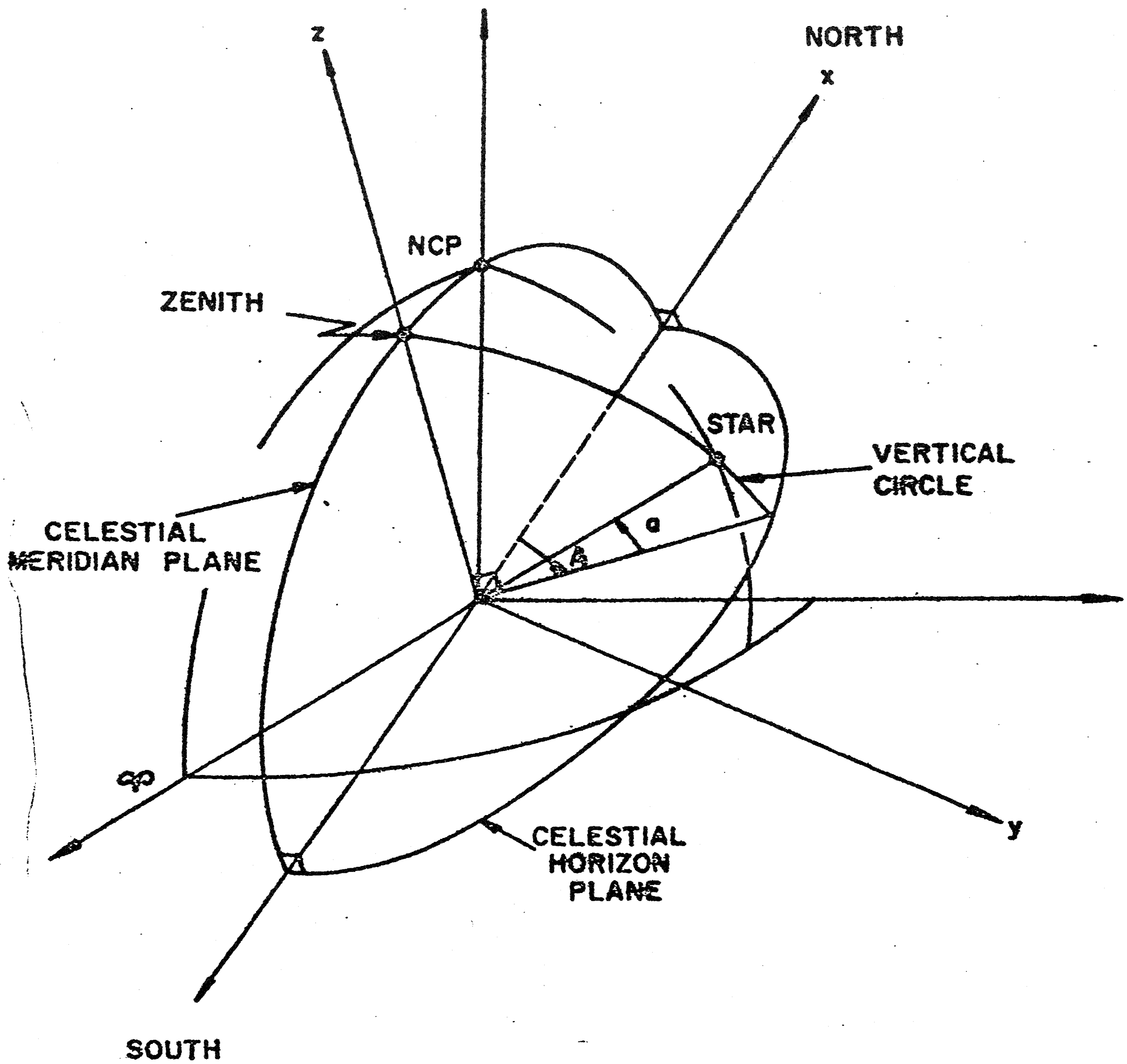


Figure 3-5.

HORIZON SYSTEM

The great circle containing the primary pole and the celestial body being observed is called the vertical circle. The location of this great circle is given by the astronomic azimuth A, the angle measured clockwise in the horizon plane from north to the vertical circle. The altitude a of the body is the angle between the horizon plane and the line directed from the origin of the system toward the body. The unit vector to a celestial body in the horizon system is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_H = \begin{bmatrix} \cos a \quad \cos A \\ \cos a \quad \sin A \\ \sin a \end{bmatrix}} . \qquad 3\text{-}11$$

The horizon system is related to the hour angle system by the astronomic latitude Φ. That is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_H = R_3\,(180^\circ)\; R_2\,(90^\circ - \Phi) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{HA}} . \qquad 3\text{-}12$$

3.5 VARIATIONS OF THE RIGHT ASCENSION SYSTEM

As mentioned at the beginning of this chapter, the celestial sphere approximation requires corrections for precise work. These corrections are for proper motion, precession, nutation, aberration, parallax and refraction, and are applied in four stages between the system in which observations are actually made (which we will call the "observed place system at epoch T") and the most absolute right ascension system (which we will call the "mean celestial system at standard epoch T_o"). We will consider these systems in the reverse order, that is:

a) mean celestial system at standard epoch T_o,

b) mean celestial system at epoch T,

c) true celestial system at epoch T,

d) apparent place system at epoch T,

e) observed place system at epoch T.

The connections between these five systems are shown in Figure 3-6. The first three systems are related by motions of the coordinate system, while the last two are related by physical effects which cause the position of the celestial body to vary.

3.5.1 Precession and Nutation

The earth is not perfectly spherical, but has an equatorial bulge which is attracted by the sun, moon, and planets in a non-symmetrical way. This causes the earth's axis of rotation (the north celestial pole) to move around the north ecliptic pole with a period of about 25,800 years and an amplitude equal to the obliquity of the ecliptic ($23\overset{\circ}{.}5$). This motion is called precession and is similar to the precession of an ordinary gyroscopic top about the gravity vector [Mueller 1969, pages 59-62].

Precession is itself not a regular motion since the earth's orbit is not circular and the moon's orbit does not lie in the ecliptic plane, and is not circular. Therefore, the added effects of the sun and moon are constantly changing as their configuration changes. Irregularities in precession are called nutation, and for the celestial pole have a period of about 18.6 years and a maximum amplitude of about 9". The added irregularity due to the changing configuration of

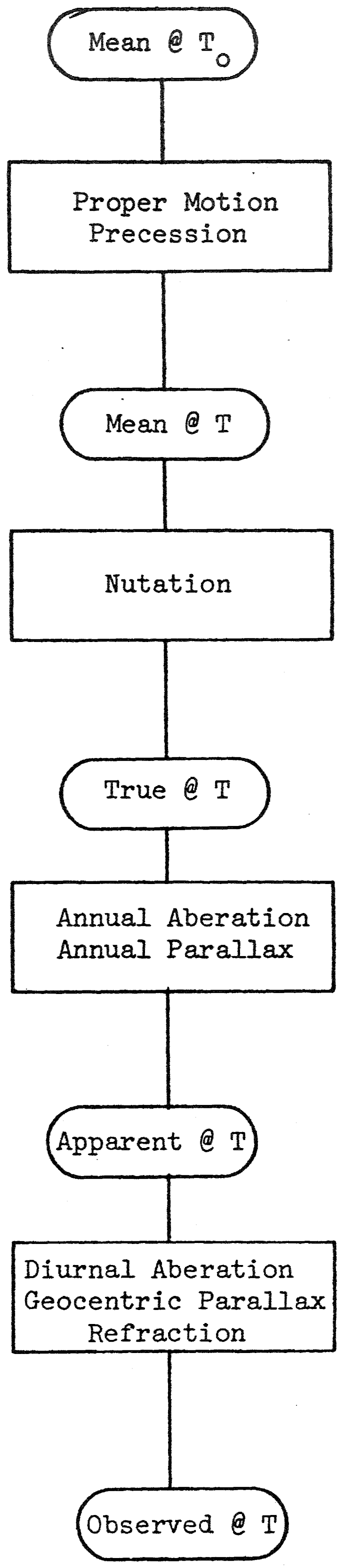


Figure 3-6.

VARIATIONS OF THE CELESTIAL RIGHT ASCENSION SYSTEM the planets is called planetary precession, and causes the very slow motion of the ecliptic plane mentioned in section 3.1. In principle, precession exists only because the earth has an equatorial bulge, and nutation and planetary precession exist only because precession exists.

Precession and nutation are shown schematically in Figure 3-7.

The celestial equator is defined as being perpendicular to the celestial pole, so that it too follows the precession and nutation of the pole. The vernal equinox is defined as being at the intersection of the celestial equator and ecliptic, so that it will follow both the precession and nutation of the celestial equator, and the motion of the ecliptic due to planetary precession. The effects are shown in Figure 3-8.

3.5.2 Mean Celestial Systems

A mean celestial (M.C.) system is specified as follows.

a) The origin is at the centre of the sun.

b) The primary pole (z-axis) is a precessing (but not nutating) pole which follows the precession of the north celestial pole, and is called the mean celestial pole.

c) The primary axis (x-axis) is a precessing (but not nutating) axis which follows the motion of the vernal equinox due both to precession of the celestial equator and rotation of the ecliptic, and is called the mean vernal equinox.

d) The y-axis is chosen so the system is right-handed.

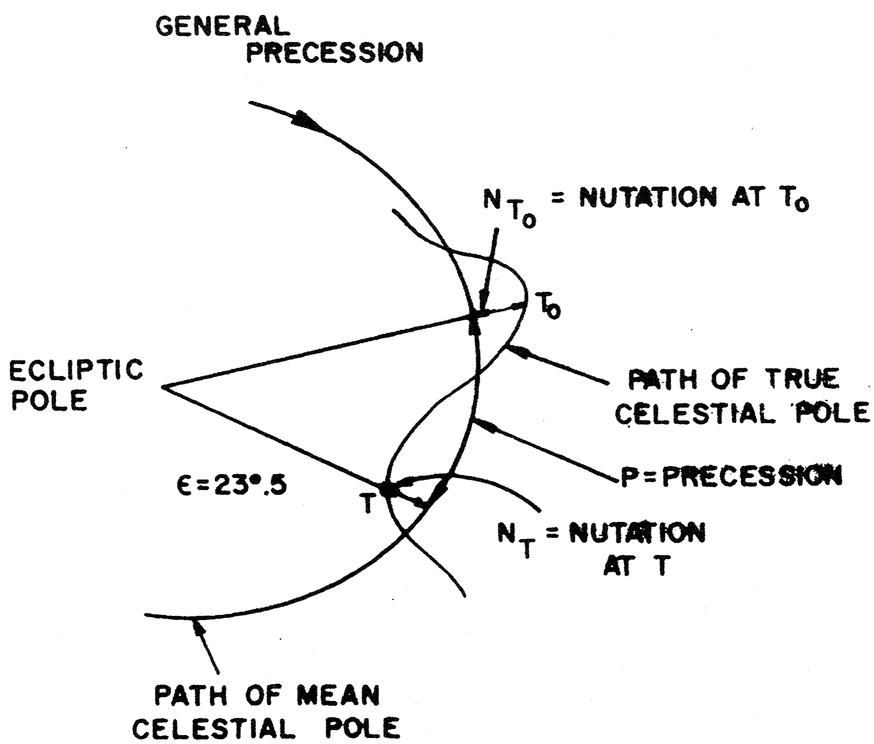


Figure 3-7. MOTION OF CELESTIAL POLE

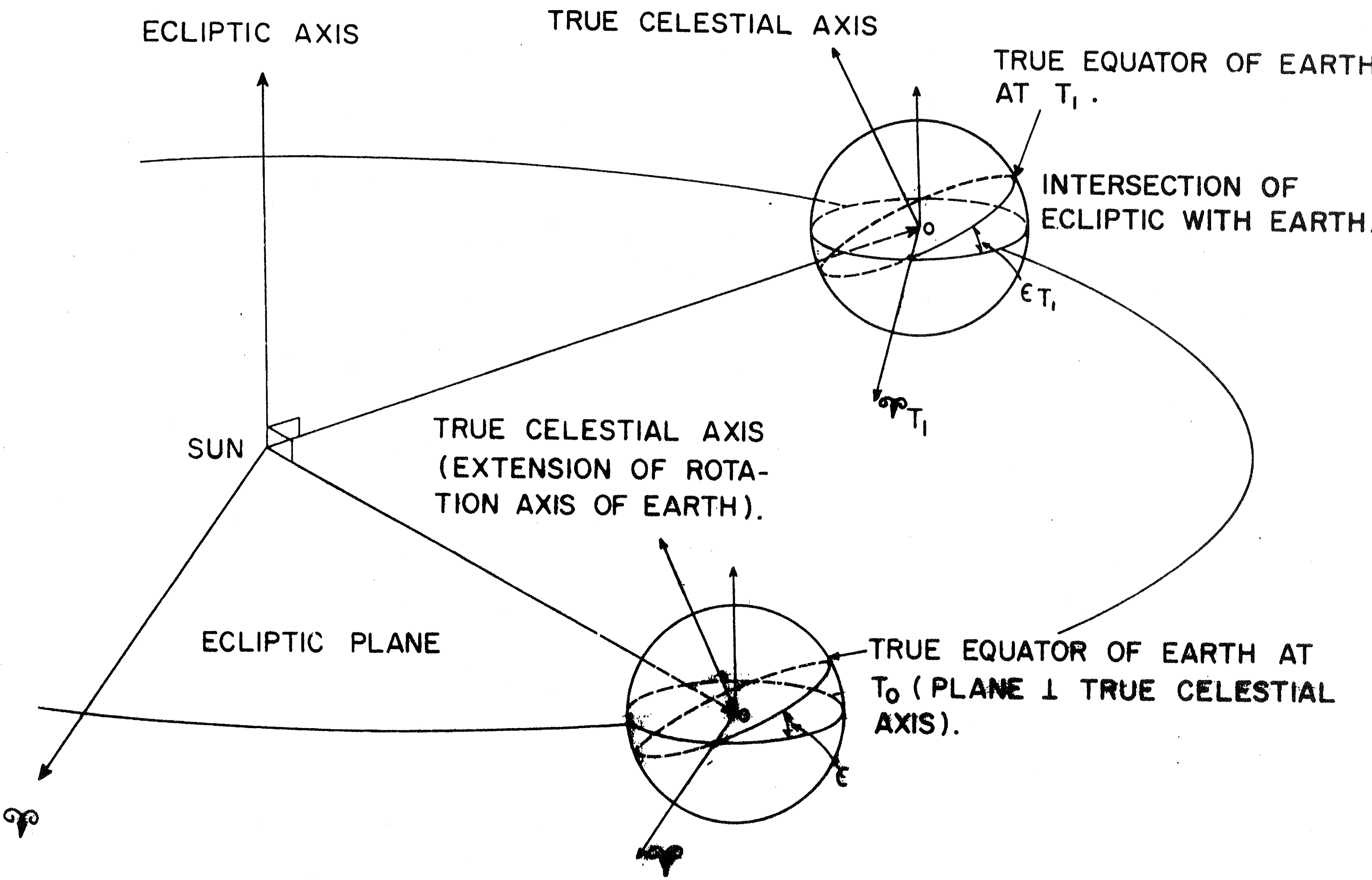
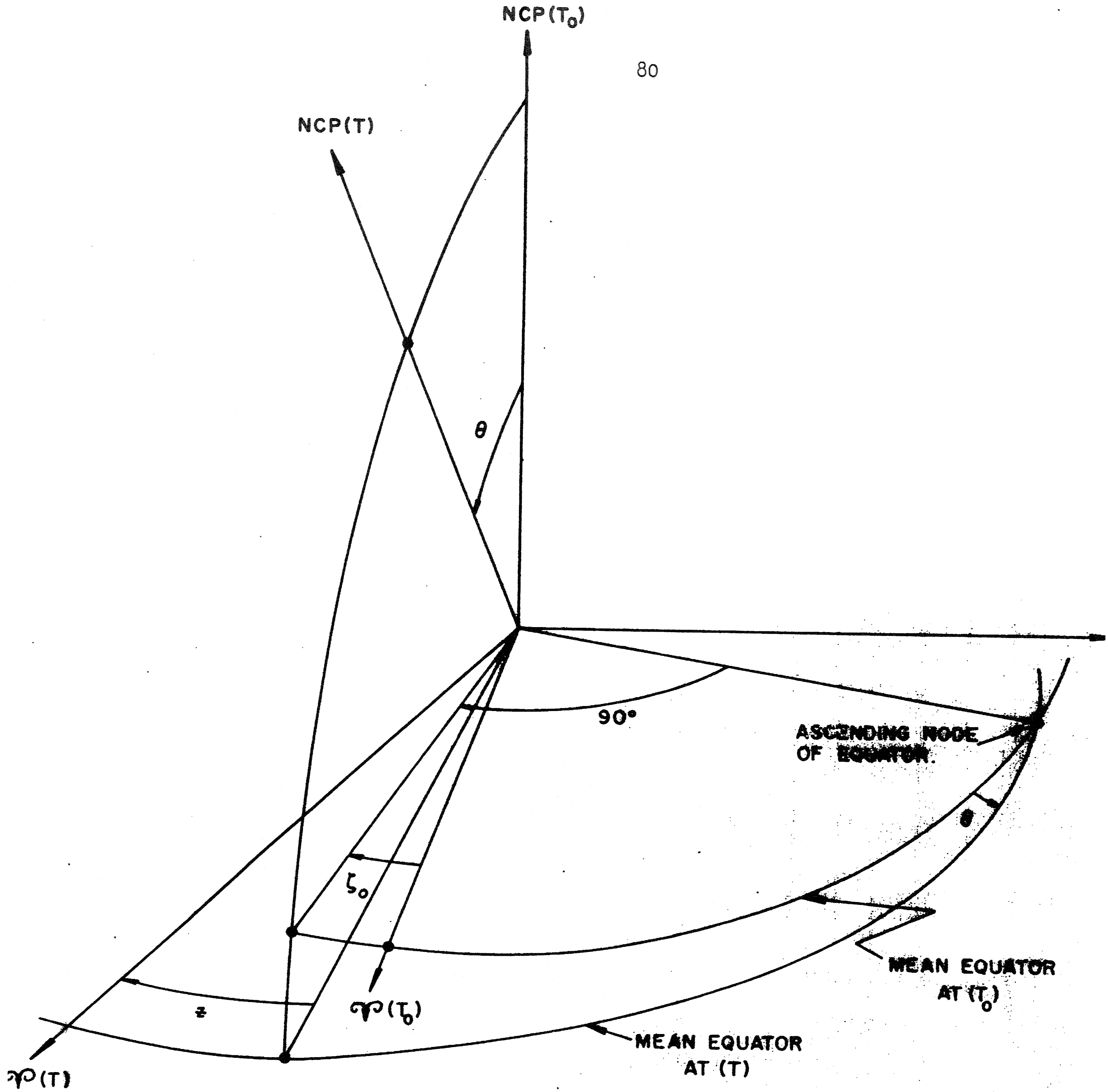


Figure 3-8. **THE EFFECT OF PRECESSION AND NUTATION**

Because the mean celestial system is moving, the coordinates (right ascension α, and declination δ) of celestial bodies vary with time. Therefore, for each epoch of time T, a different mean celestial system is defined. Certain epochs T_o have been chosen as <u>standard epochs</u>, to which tabulated mean celestial coordinates of celestial bodies refer. The relationship between mean celestial systems of times T_o and T is usually defined in terms of the <u>precessional elements</u> (ζ_o, θ, z) as shown in Figure 3-9. Expressions for these elements as a function of time were derived over 70 years ago by Simon Newcomb [Mueller 1969, p. 63]. The angles ($90°-\zeta_o$) and ($90°+z$) are the right ascensions of the ascending node of the equator at T measured respectively in the systems at T_o and at T. The angle θ is the inclination of the equator at T with respect to the equator at T_o. The transformation from a mean celestial system at T_o to one at T is given by

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{M.C.T.} = R_3(-z)\ R_2(\theta)\ R_3(-\zeta_o) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{M.C.T_o} .} \qquad 3\text{-}13$$

Independent of the motion of the mean celestial coordinate system due to precession, each star is changing in position due to proper motion. Because this proper motion is uniform, it is most appropriate to account for it in the most uniform right ascension system, that is the mean celestial system. The proper motion components for each star of interest (usually tabulated



PRECESSION

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{M.C._T} = R_3(-z)\, R_2(\theta)\, R_3(-\zeta_o) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{M.C._{T_o}}$$

Figure 3-9. MEAN CELESTIAL COORDINATE SYSTEMS as rates of changes in right ascension and declination) must therefore be included in the conversion of mean place at T_o to mean place at T.

3.5.3 The True Celestial System

A true celestial system (T.C.) is specified as follows.

a) The origin is at the centre of the sun.

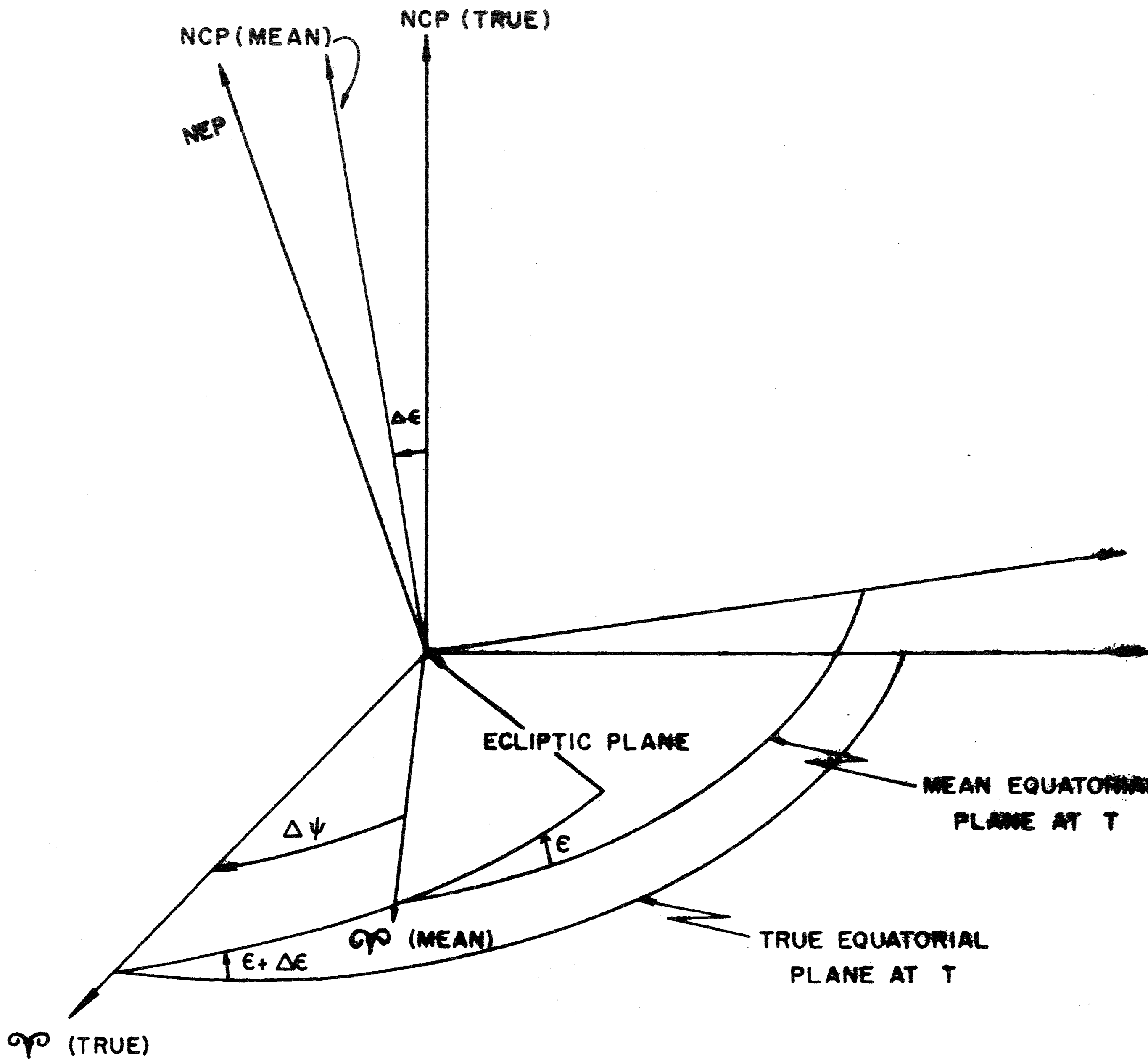
b) The primary pole (z-axis) is a precessing and nutating pole which follows the precession and nutation of the north celestial pole, and is called the true celestial pole.

c) The primary axis (x-axis) is a precessing and nutating axis which follows the motion of the vernal equinox due to precession and nutation of the celestial equator, and to rotation of the ecliptic, and is called the true vernal equinox.

d) The y-axis is chosen so the system is right-handed.

As in the case of mean celestial systems, a different true celestial system is defined for each epoch of time T. The true celestial system at epoch T differs from the mean celestial system at epoch T only by the effect of nutation, and the relationship is usually defined in terms of the nutation in longitude $\Delta\psi$ and nutation in obliquity $\Delta\varepsilon$ shown in Figure 3-10. Expressions for these two elements as a function of time, and other parameters were derived by Woolard [Mueller 1969, p. 69]. The transformation from a mean celestial system at T to a true celestial system at T is given by

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{T.C.T.} = R_1(-\varepsilon-\Delta\varepsilon)\ R_3(-\Delta\psi)\ R_1(\varepsilon) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{M.C.T.} .} \qquad 3\text{-}14$$



NUTATION

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{T.C._T} = R_1(-\epsilon-\Delta\epsilon)\, R_3(-\Delta\psi)\, R_1(\epsilon) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{M.C._T}$$

TRUE AND MEAN CELESTIAL COORDINATE SYSTEMS.

Figure 3-10.

3.5.4 The Apparent Place System

An apparent place (A.P.) system is specified as follows.

a) The origin is at the centre of the earth.

b) The primary pole is parallel to the true celestial pole.

c) The primary axis is parallel to the true vernal equinox.

d) The system is right-handed.

Therefore, an apparent place system is a true celestial system with the origin shifted from the centre of the sun to the centre of the earth. This means the origin is no longer at the centre of the true celestial sphere which causes annual parallax, and the origin is revolving around the centre of the true celestial sphere which causes annual aberration.

If the earth's orbit is regarded as circular, the earth has a constant of aberration.

$$\kappa = \frac{v}{c} \text{ cosec } 1'' = 20''.4958 \quad , \qquad \text{3-15}$$

where v is the earth's velocity and c the velocity of light; and the radius of the earth's orbit will subtend a different angle Π at each star, called the stellar parallax for that star. The nearest star has a stellar parallax of 0".76.

The right ascension α and declination δ of a star expressed in the apparent place system is then [Mueller 1969, pages 93 and 61].

$$\boxed{\begin{bmatrix} \alpha \\ \delta \end{bmatrix}_{AP} = \begin{bmatrix} \alpha \\ \delta \end{bmatrix}_{TRUE} + \begin{bmatrix} \Delta\alpha_P \\ \Delta\delta_P \end{bmatrix} + \begin{bmatrix} \Delta\alpha_A \\ \Delta\delta_A \end{bmatrix} ,} \qquad \text{3-16}$$

where

$$\begin{bmatrix} \Delta\alpha_p \\ \Delta\delta_p \end{bmatrix} = \Pi \begin{bmatrix} \cos\alpha \cos\varepsilon \sec\delta - \sin\alpha \cos\lambda_s \sec\delta \\ \cos\delta \sin\varepsilon \sin\lambda_s - \cos\alpha \sin\delta \cos\lambda_s - \sin\alpha \sin\delta \cos\varepsilon \sin\lambda_s \end{bmatrix} \quad 3\text{-}17$$

and

$$\begin{bmatrix} \Delta\alpha_A \\ \Delta\delta_A \end{bmatrix} = -\kappa \begin{bmatrix} \cos\alpha \cos\lambda_s \cos\varepsilon \sec\delta + \sin\alpha \sin\lambda_s \sec\delta \\ \cos\lambda_s \cos\varepsilon(\tan\varepsilon \cos\delta - \sin\alpha \sin\delta) + \cos\alpha \sin\delta \sin\lambda_s \end{bmatrix} \quad 3\text{-}18$$

and λ_s is the longitude of the sun, ε the obliquity of the ecliptic, and (α, δ) in 3-17 and 3-18 expressed in the true celestial system.

The fact that the earth's orbit is not circular introduces errors of about 1% in equation 3-17 and up to $0\overset{''}{.}343$ in equation 3-18.

3.5.5 The Observed Place System

An observed place (O.P.) system is specified as follows.

a) The origin is at the observing station.

b) The primary pole is parallel to the true celestial pole.

c) The primary axis is parallel to the true vernal equinox.

d) The system is right-handed.

Therefore, an observed place system is an apparent place system with the origin shifted from the centre of the earth to the observing station. This means the origin is no longer at the centre of the earth, which causes <u>geocentric parallax</u>, and the origin is rotating around the centre of the earth, which causes <u>diurnal aberration</u>. In fact, the effect of geocentric parallax is always negligible when observing stars. The <u>diurnal constant of aberration</u> is

$$k = \frac{v}{c} \operatorname{cosec} 1'' = 0\overset{''}{.}320\ \rho \cos\phi \qquad 3\text{-}19$$

where v is the earth's surface rotational velocity, c is the velocity of light, ρ is the radial distance from geocenter to observer in units of earth radius, and ϕ is the geodetic latitude of the observer.

There is a third effect due to the fact that the earth is blanketed with an atmosphere of varying optical density. This causes a complex change in the direction of the light ray from a star which depends on the incident angle. Mueller [1969, pages 103-109] discusses this atmospheric refraction in detail.

The right ascension and declination of a star in the observed place system is then

$$\boxed{\begin{bmatrix} \alpha \\ \delta \end{bmatrix}_{O.P.} = \begin{bmatrix} \alpha \\ \delta \end{bmatrix}_{A.P.} + \begin{bmatrix} \Delta\alpha_D \\ \Delta\delta_D \end{bmatrix} - \begin{bmatrix} \Delta\alpha_R \\ \Delta\delta_R \end{bmatrix},} \qquad 3\text{-}20$$

where

$$\begin{bmatrix} \Delta\alpha_D \\ \Delta\delta_D \end{bmatrix} = k \begin{bmatrix} \cos h & \sec\delta \\ \sin h & \sin\delta \end{bmatrix} \qquad 3\text{-}21$$

where h is the hour angle of the star, and $(\Delta\alpha_R, \Delta\delta_R)$ are the corrections due to refraction.

3.6 TRANSFORMATION BETWEEN APPARENT CELESTIAL AND AVERAGE TERRESTRIAL COORDINATE SYSTEMS

The apparent celestial and average terrestrial coordinate systems both have

a) their origins at the centre of gravity of the earth,

b) their primary poles as the CIO pole, that is the average terrestrial pole is parallel to the true celestial pole,

c) both are right-handed.

The only difference between the two systems is that the primary axis of the apparent celestial system is parallel to the true vernal equinox, and the primary axis of the average terrestrial system lies in the Greenwich mean astronomic meridian. The angle between these two axes varies with the rotation of the earth, and is called the Greenwich Apparent Sidereal Time (GAST). Therefore, the transformation from apparent celestial to average terrestrial is (see Figure 3-4).

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = R_3\ (GAST) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.P.} .} \qquad 3\text{-}22$$

To use this equation, we require some means of computing GAST from the Universal (Solar) Time used for broadcasts of standard time. We will describe two methods.

First, if GAST is known for some epoch T_o of Universal Time, then it may be computed for some other epoch T from the relation

$$\boxed{GAST\ (T) = GAST\ (T_o) + \omega_e(T-T_o),} \qquad 3\text{-}23$$

where it is assumed that sidereal and universal time are related by a uniform rotation rate of the earth

$$\begin{aligned} \omega_e &= 360.98565 \text{ degrees/UT day} \\ &= 4.3752695 \times 10^{-3} \text{ radians/minute.} \end{aligned} \qquad 3\text{-}24$$

This is not precisely true, but a difference with respect to the more accurate method presented below of less than 10^{-7} radians (equivalent to about 0.02 arcseconds, 1 millisecond, or 1/2 meter along the earth's equator) is introduced if $(T-T_o)$ is less than a day.

A more accurate relation is given by Veis [1966, p. 19]:

$$\boxed{\begin{aligned} GAST = {} & 100°.075542 + 360°.985647348\ T + 0°.2900 \times 10^{-12}\ T^2 \\ & - 4°.392 \times 10^{-3} \sin (12°.1128 - 0°.052954\ T) \\ & + 0°.053 \times 10^{-3} \sin 2(12°.1128 - 0°.052954\ T) \\ & - 0°.325 \times 10^{-3} \sin 2(280°.0812 + 0°.9856473\ T) \\ & - 0°.050 \times 10^{-3} \sin 2(64°.3824 + 13°.176398\ T)\ , \end{aligned}} \qquad 3\text{-}25$$

where T is the number of Julian Days since the epoch 0.5 January 1950 (that is midnight of December 31, 1949). For 1971

$$T = 7669 + D + (M + S/60)/1440\ , \qquad 3\text{-}26$$

where D = the day number during 1971,

M = the minute of UT time,

S = the second of UT time,

and 7669 is the number of days between January 1, 1950 and December 31, 1970. This expression is accurate to 0.2 arcseconds, 10 milliseconds, or 5 meters along the equator for any value $(T - T_o)$. More accuracy can be obtained by adding more terms [Nautical Almanac Office 1961].

3.7 SUMMARY OF CELESTIAL SYSTEMS

In this chapter we have defined four celestial coordinate systems:

a) Ecliptic (E),

b) Right Ascension (R.A.),

c) Hour Angle (H.A.),

d) Horizon (H.).

Table 3-1 summarizes the reference poles, planes and axes defining the systems. Table 3-2 summarizes the transformations between these systems.

We have also precisely defined four variations to the right ascension system:

a) Mean Celestial (M.C.),

b) True Celestial (T.C.),

c) Apparent Place (A.P.),

d) Observed Place (O.P.),

all of which vary with time, so that the epoch T to which they refer must be specified. Figure 3-11 shows the parameters which connect all of these celestial coordinate systems.

Table 3-1

REFERENCE POLES, PLANES AND AXES DEFINING CELESTIAL COORDINATE SYSTEMS

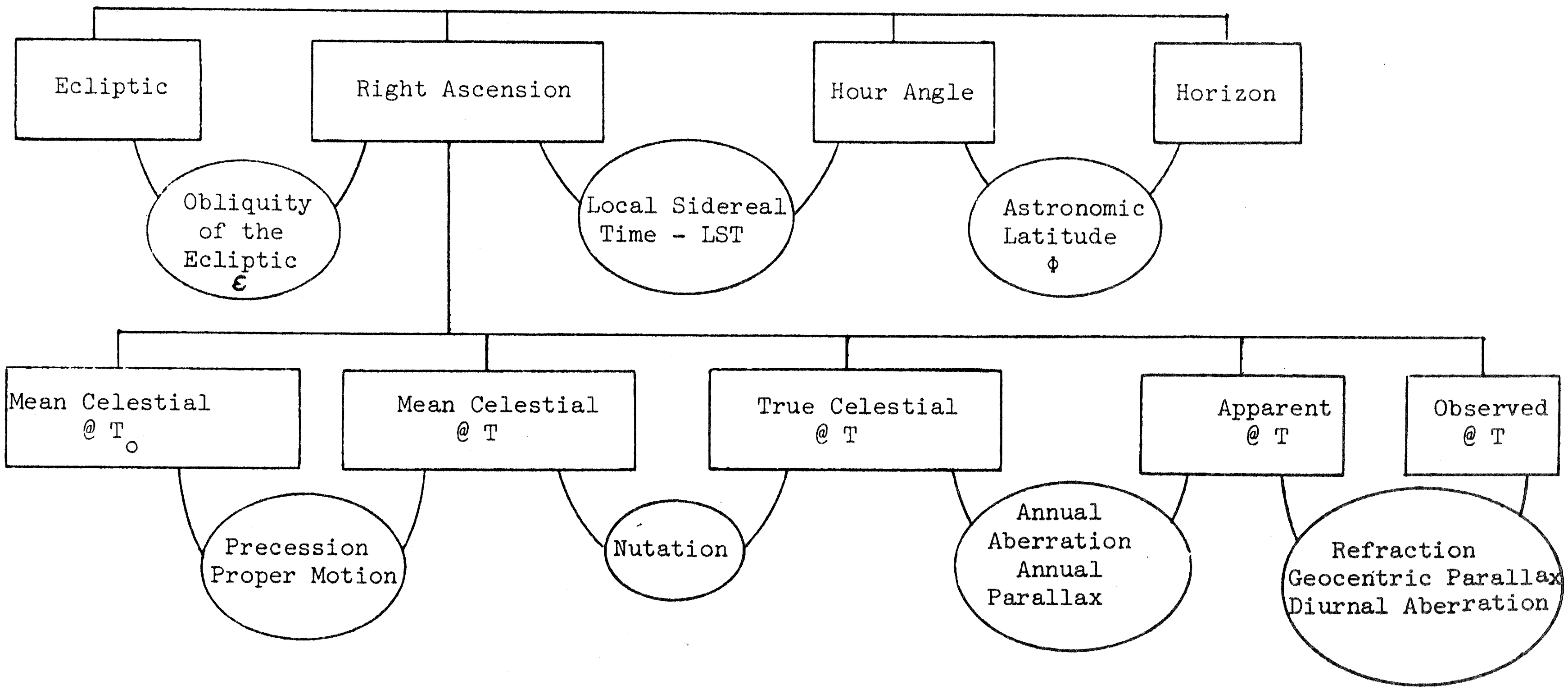
System	Reference Poles		Reference Planes		
	Primary pole (z-axis)	Secondary Pole (x-axis)	Primary	Secondary	Handedness (y-axis)
Ecliptic	North ecliptic pole	Vernal equinox	Ecliptic	Ecliptic meridian of of the equinox (half containing vernal equinox)	right
Right ascension	North celestial pole	Vernal equinox	Celestial equator	Equinoctial colure (half containing vernal equinox)	right
Hour angle	North celestial pole		Celestial equator	Hour circle of observer's zenith (half containing zenith)	left
Horizon	Zenith	North point	Celestial horizon	Celestial meridian (half containing north pole)	left

Table 3-2

TRANSFORMATIONS AMONG CELESTIAL COORDINATE SYSTEMS.

		Original System			
		Ecliptic	Right Ascension	Hour Angle	Horizon
Final System	Ecliptic	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_E$	$R_1\ (\epsilon)$		
	Right Ascension	$R_1\ (-\epsilon)$	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{R.A.}$	$R_3\ (-LST)\ P_2$	
	Hour Angle		$P_2\ R_3(+LST)$	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{H.A.}$	$R_2(\Phi-90°)\ R_3(180°)$
	Horizon			$R_3(180°)\ R_2(90°-\Phi)$	$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{H.}$

Figure 3-11. Celestial Coordinate Systems.



4. THE ORBITAL COORDINATE SYSTEM

In this chapter we discuss the orbital system, which is used to define the coordinates of a satellite orbiting around the earth. We first discuss the orbital ellipse, and the coordinate system in the orbit plane. Then we transform this system into the apparent celestial and average terrestrial systems, and discuss variations in the orbital elements. Finally expressions for the coordinates of the satellite subpoint, and the topocentric coordinates of the satellite are derived.

4.1 THE ORBITAL ELLIPSE AND ORBITAL ANOMALIES

The trajectory of a body moving in a central force field describes an ellipse, with the attracting force centred at one of the foci of the ellipse.

In the case of a satellite orbiting around the earth, this is called the orbital ellipse, and the centre of gravity of the earth is at one of the foci (see Figure 4-1). The point of closest approach of the satellite to the earth is called the perigee, and the farthest point is called the apogee. Both perigee and apogee lie on the semi-major axis of the ellipse, called the line of apsides. The size and shape of the orbital ellipse are usually defined using the semi-major axis, a and the eccentricity e, where

$$e^2 = \frac{a^2 - b^2}{a^2} \qquad 3\text{-}1$$

and b is the semi-minor axis of the ellipse.

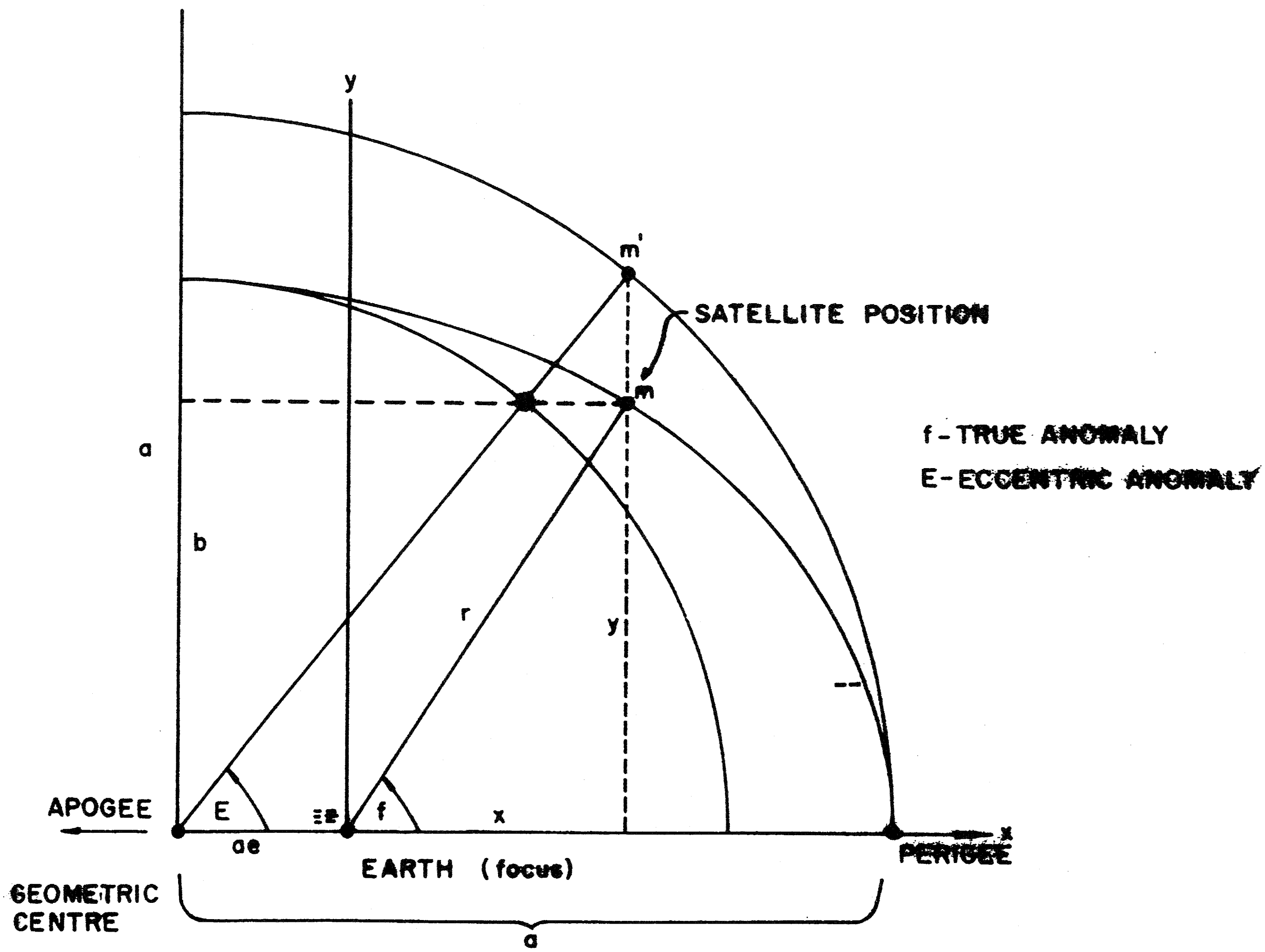


FIGURE 4-1 **ORBIT ELLIPSE**

Consider the satellite to be at a point m on the orbital ellipse. The angular distance between perigee and m is called the satellite anomaly. There are three anomalies. The true anomaly f is the angle between the line of apsides and the line joining the focus to the satellite.

Consider the projection of the satellite position m along a line parallel to the semi-minor axis to intersect a circle with radius equal to the semi-major axis at a point m'. The eccentric anomaly E is the angle between the line of apsides and the line joining the geometric centre of the ellipse to m'.

The mean anomaly $\bar{M}$ is the true anomaly corresponding to the motion of an imaginary satellite of uniform angular velocity, that is $\bar{M} = 0$ at the perigee and then increases uniformly at a rate of 360° per revolution. When this is expressed as a rate per unit time, then it is called the mean anomalistic motion $\bar{n}$.

The relationship between the true anomaly f and the eccentric anomaly E is from Figure 4-1

$$\boxed{\begin{bmatrix} x \\ y \end{bmatrix} = r \begin{bmatrix} \cos f \\ \sin f \end{bmatrix} = \begin{bmatrix} a \cos E - ae \\ b \sin E \end{bmatrix} = \begin{bmatrix} a(\cos E - e) \\ a(1 - e^2)^{1/2} \sin E \end{bmatrix},} \qquad 4\text{-}2$$

or

$$\boxed{\tan f = \frac{(1 - e^2)^{1/2} \sin E}{\cos E - e} .} \qquad 4\text{-}3$$

The relationship between the eccentric anomaly E and the mean anomaly $\bar{M}$ is Kepler's equation [Kaula 1966, p. 23]

$$\boxed{\bar{M} = E - e \sin E.} \qquad 4\text{-}4$$

where $\bar{M}$ and E are in radians.

We are usually given the mean anomaly $\bar{M}$, and want to find the eccentric anomaly E from equation 4-4. We will present three ways.

If the eccentricity is very small (say $e \approx 0.002$), then the $e \sin E$ term will be small, and $\bar{M} \approx E$. Therefore we can write

$$\boxed{E = \bar{M} + e \sin E \approx \bar{M} + e \sin \bar{M} \quad .} \tag{4-5}$$

For an eccentricity of $e = 0.002$ this approximation introduces an error of about 10^{-6} radians.

If greater precision is required or the eccentricity is not so small, we can solve 4-4 iteratively. Taking the total differential of 4-4

$$\delta M = (1 - e \cos E)\, \delta E$$

or

$$\delta E = \frac{\delta \bar{M}}{1 - e \cos E} \quad . \tag{4-6}$$

Given $\bar{M}$, the iterative solution of 4-4 begins by making an initial approximation from 4-5 as

$$E_o = \bar{M} + e \sin \bar{M} \; .$$

The following equations are then iteratively evaluated in order

$$\boxed{\begin{aligned} \bar{M}_i &= E_i - e \sin E_i \\ \Delta\bar{M} &= \bar{M}_i - \bar{M} \\ \Delta E &= \frac{\Delta\bar{M}}{1 - e \cos E_i} \\ E_{i+1} &= E_i + \Delta E \end{aligned}}$$

until the difference $\Delta\bar{M}$ is less than some chosen ε.

A third method of evaluating E is to use a power series in e, for example [Brouwer and Clemence 1961, p. 76]:

$$\boxed{\begin{aligned} E = \bar{M} &+ \left(e - \frac{1}{8} e^3 + \frac{1}{192} e^5 - \frac{1}{9216} e^7\right) \sin \bar{M} + \\ &+ \left(\frac{1}{2} e^2 - \frac{1}{6} e^4 + \frac{1}{98} e^6\right) \sin 2\bar{M} + \\ &+ \left(\frac{3}{8} e^3 - \frac{27}{128} e^5 + \frac{243}{5120} e^7\right) \sin 3\bar{M} + \\ &+ \left(\frac{1}{3} e^4 - \frac{4}{15} e^6\right) \sin 4\bar{M} + \left(\frac{125}{384} e^5 - \frac{3125}{9216} e^7\right) \sin 5\bar{M} + \\ &+ \frac{27}{80} e^6 \sin 6\bar{M} + \frac{16807}{46080} e^7 \sin 7\bar{M} \,. \end{aligned}} \qquad 4\text{-}7$$

4.2 THE ORBITAL COORDINATE SYSTEM

The orbital (ORB) coordinate system is specified as follows (see Figure 4-2):

a) The origin is at the centre of gravity of the earth.

b) The primary plane is the plane of the orbital ellipse, and the primary pole (z - axis) is perpendicular to this plane (see Figure 4-1).

c) The primary axis (x - axis) is the line of apsides.

d) The y-axis is chosen so that the system is right-handed.

The position vector of the satellite in its orbit is given by

$$\boxed{\bar{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{ORB} = r \begin{bmatrix} \cos f \\ \sin f \\ 0 \end{bmatrix} = \begin{bmatrix} a\,(\cos E - e) \\ a(1 - e^2)^{1/2} \sin E \\ 0 \end{bmatrix} \,.} \qquad 4\text{-}8$$

Note z = 0 because the satellite is assumed not to be out of the orbit plane.

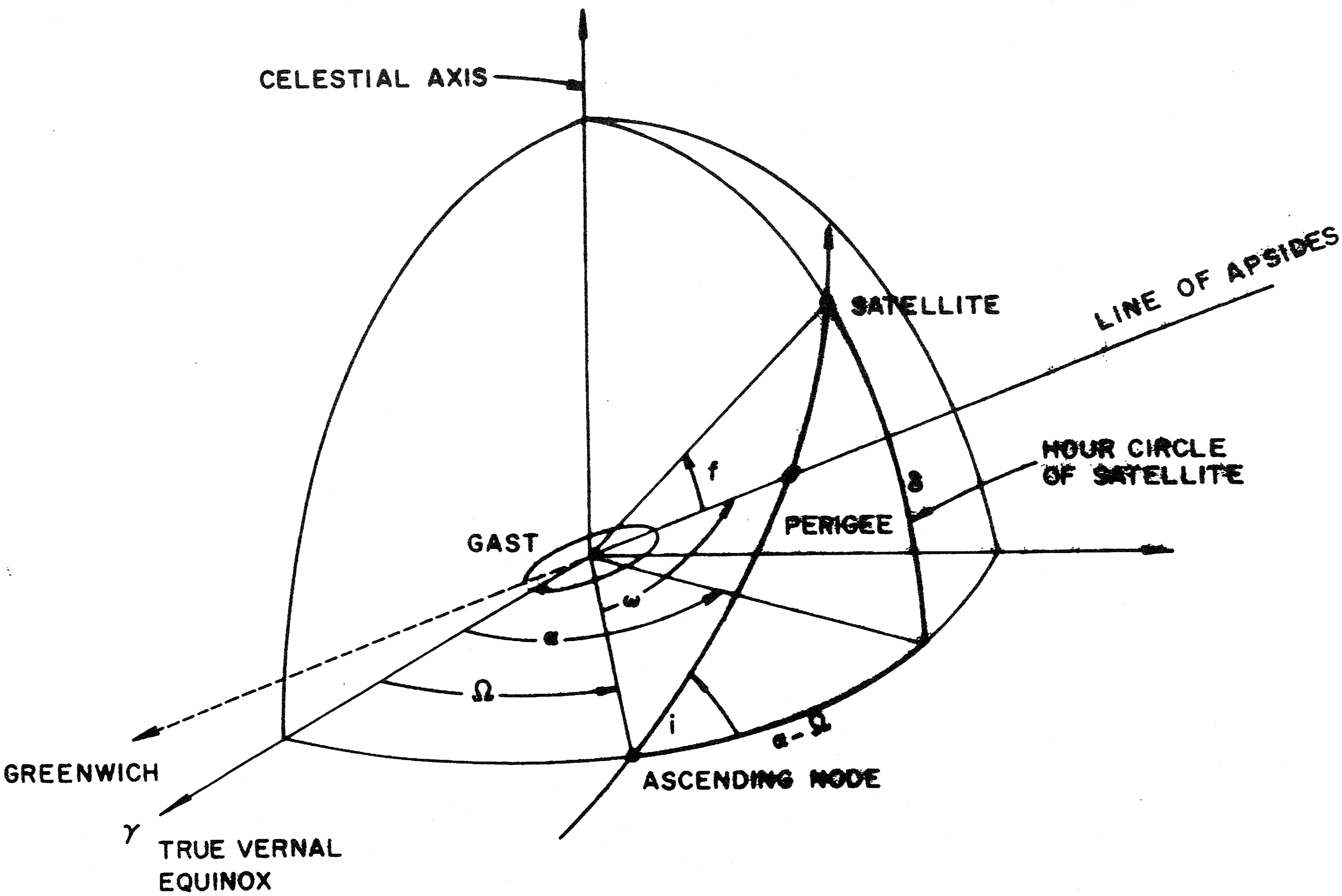


FIGURE 4-2 KEPLERIAN ORBITAL ELEMENTS

a - SEMI MAJOR, e - ECCENTRICITY: SIZE AND SHAPE OF ORBIT

ω - ARGUMENT OF PERIGEE - DIRECTION OF LINE OF APSIDES, Ω - RIGHT ASCENSION OF ASCENDING NODE, i - INCLINATION: POSITION OF ORBIT

f - TRUE ANOMALY AT TIME T

4.3 TRANSFORMATION FROM ORBITAL TO AVERAGE TERRESTRIAL SYSTEM

The orbital plane does not rotate with the earth, but remains fixed in the celestial system. The orbital system and the apparent celestial both have their origins at the centre of gravity of the earth.

From Figure 4-2 we see that when the orbital plane is extended to meet the celestial sphere, it intersects the celestial equator at the <u>ascending node</u> (where the satellite croses the equator from south to north), and the <u>descending node</u>. The angle between the celestial equator and the orbital plane is the <u>inclination</u> i. The angle between the ascending node and the line of apsides, measured in the orbital plane is the <u>argument of pergiee</u> ω. The angle between the vernal equinox and the ascending node, measured in the celestial equatorial plane is the <u>right ascension of the ascending node</u> Ω.

The transformation from the orbital system to the apparent celestial system is

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.P.} = R_3(-\Omega)\ R_1(-i)\ R_3(-\omega) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{ORB} .} \qquad 4\text{-}9$$

The transformation from apparent celestial to average terrestrial is given by equation 3-22, so that

$$\boxed{\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = R_3(GAST)\ R_3(-\Omega)\ R_1(-i)\ R_3(-\omega) \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{ORB} .} \qquad 4\text{-}10$$

4.4 VARIATIONS IN THE ORBITAL ELEMENTS

So far we have assumed that the satellite orbit does not vary with time, and is completely specified by the six Keplerian orbital elements a, e, f, ω, i, Ω. The earth's gravitational force field is not spherically symmetric, as evidenced by geoid undulations and the equatorial bulge. Also the atmosphere exerts a fluctuating drag force on the satellite. Because of these and other smaller effects, the satellite trajectory cannot be assumed to be a fixed ellipse. However, for each epoch of time T there will be a different orbital ellipse tangent to the satellite trajectory at that epoch. Each of these different ellipses will have a set of Keplerian orbital elements, and if the variation in these orbital elements with time is known, they are said to describe an osculating orbital ellipse which describes the satellite trajectory accurately.

The variation with time of the inclination angle is equivalent to introducing a time-varying out-of-plane component, and this is often done. For nearly circular orbits, the eccentricity is small to start with, so variations are usually neglected.

4.5 THE SATELLITE SUBPOINT

The subpoint of a satellite is simply the trace of the path of the satellite on the ellipsoid (see Figure 4-3). The coordinates of the subpoint are given by the geodetic latitude ϕ and longitude λ of the ellipsoidal normal passing through the satellite.

The average terrestrial Cartesian coordinates of the satellite are known from equation 4-10. From equation 2-32

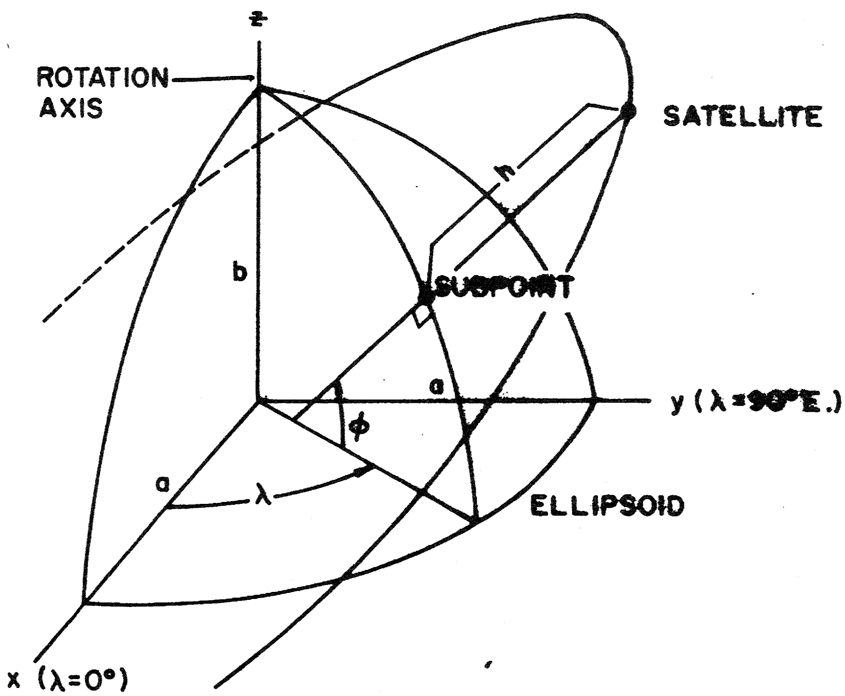


Figure 4-3 **SATELLITE SUBPOINT**

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{A.T.} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} + \begin{bmatrix} (N + h) \cos\phi \cos\lambda \\ (N + h) \cos\phi \sin\lambda \\ (Nb^2/a^2 + h) \sin\phi \end{bmatrix} ,$$

and if the reference ellipsoid parameters (a, b, x_o, y_o, z_o) are known this equation can be iteratively inverted to solve for (ϕ, λ, h) using the method of section 2.2.6.

4.6 TOPOCENTRIC COORDINATES OF SATELLITE

If we are observing a satellite at position j from a station i on the earth (see Figure 4-4), then we will require an expression for the coordinates of the satellite in the local geodetic system of station i.

If the coordinates (ϕ_i, λ_i, h_i) of station i are known with respect to a reference ellipsoid (a, b, x_o, y_o, z_o) then the geodetic Cartesian coordinates of i

$$\bar{r}_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}_G \qquad 4\text{-}11$$

can be computed from equation 2-31.

If the average terrestrial Cartesian coordinates of j have been computed by the methods outlined in this chapter, then the geodetic coordinates of j are

$$\bar{r}_j = \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}_{A.T.} - \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} . \qquad 4\text{-}12$$

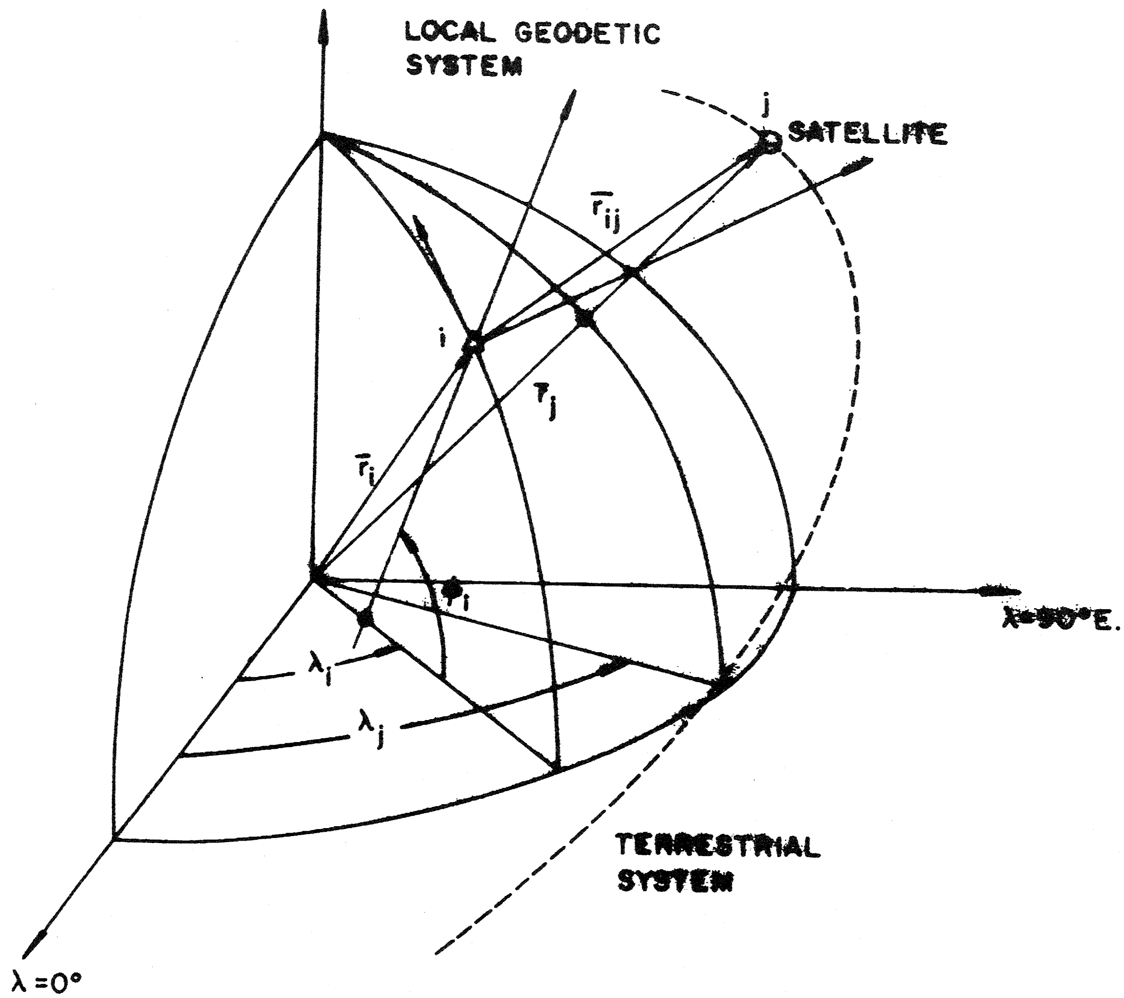


Figure 4-4 **TOPOCENTRIC COORDINATES OF SATELLITE**

The range vector from i to j is

$$\bar{r}_{ij} = \bar{r}_j - \bar{r}_i = \begin{bmatrix} x_j - x_i \\ y_j - y_i \\ z_j - z_i \end{bmatrix}_{G.} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}_G \quad . \qquad 4\text{-}13$$

The coordinates of the range vector can be expressed in the local geodetic system by using equation 2-67

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}_{L.G.} = P_2 \, R_2(\phi - 90°) \, R_3(\lambda - 180°) \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}_{G.} \quad . \qquad 4\text{-}14$$

But from equation 2-66

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}_{L.G.} = \Delta r \begin{bmatrix} \cos a \cos \alpha \\ \cos a \sin \alpha \\ \sin a \end{bmatrix} , \qquad 4\text{-}15$$

where the range

$$\Delta r = (\Delta x^2 + \Delta y^2 + \Delta z^2)^{1/2} \qquad 4\text{-}16$$

and the altitude a and azimuth α are given by

$$\boxed{\begin{aligned} a &= \sin^{-1} \left(\frac{\Delta z}{\Delta r}\right) \qquad 4\text{-}17 \\ \alpha &= \tan^{-1} \left(\frac{\Delta y}{\Delta x}\right) \quad . \qquad 4\text{-}18 \end{aligned}}$$

5. SUMMARY OF COORDINATE SYSTEMS

In this chapter we will summarize the relationships between the coordinate systems dealt with in these notes. We will also explain a duality paradox which has arisen earlier in the notes. This chapter is, in effect, an explanation of the symbols and abbreviations used in Figure 5-1.

5.1 TERRESTRIAL SYSTEMS

We discussed five terrestrial systems:

a) I.T. = Instantaneous Terrestrial Coordinate System,

b) A.T. = Average Terrestrial Coordinate System,

c) G. = Geodetic Coordinate System,

d) L.G. = Local Geodetic Coordinate System,

e) L.A. = Local Astronomic Coordinate System,

which are related to each other by the four sets of parameters

a) Polar Motion (x_p, y_p) - relates I.T. and A.T.,

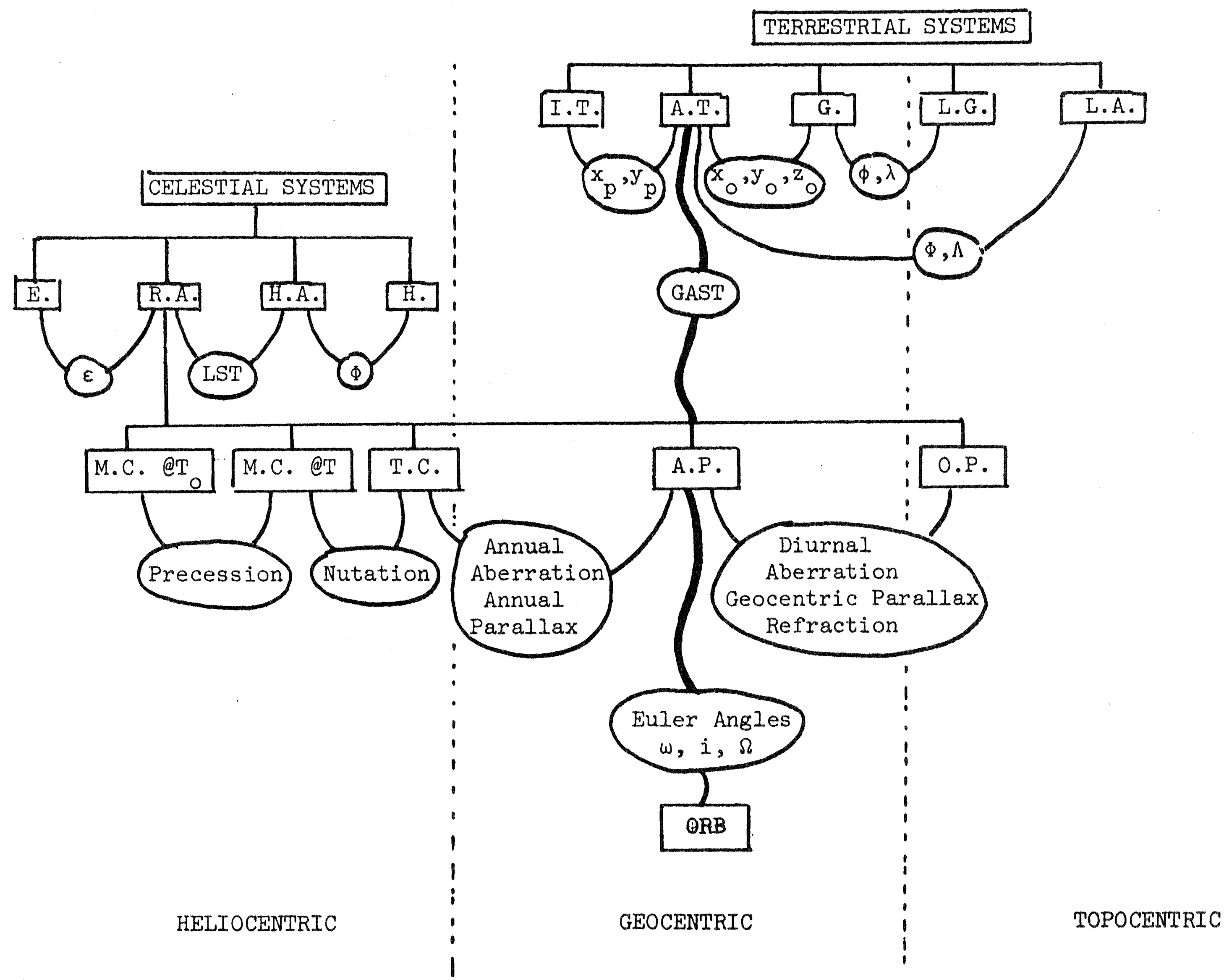
b) Translation of the origin (x_o, y_o, z_o) - relates A.T. and G.,

c) Geodetic Latitude and Longitude (ϕ, λ) - relates G. and L.G.,

d) Astronomic Latitude and Longitude (Φ, Λ) - relates A.T. and L.A.

Figure 5-1

COORDINATE SYSTEMS.



5.2 CELESTIAL SYSTEMS

We discussed four main celestial systems

a) E. = Celestial Ecliptic Coordinate System,

b) R.A. = Celestial Right Ascension Coordinate System,

c) H.A. = Celestial Hour Angle Coordinate System,

d) H. = Celestial Horizon Coordinate System,

which are related to each other by the three parameters

a) Obliquity of the ecliptic (ε) - relates E. and R.A.,,

b) Local Sidereal Time (LST) - relates R.A. and H.A.,

c) Astronomic Latitude (Φ) - relates H.A. and H.

The R.A. system has four variations

a) M.C. = Mean Celestial Coordinate System,

b) T.C. = True Celestial Coordinate System,

c) A.P. = Apparent Place Coordinate System,

d) O.P. = Observed Place Coordinate System.

which all vary with time and thus are defined only when the epoch T to which they refer is specified.

The parameters relating these systems are

a) Precession and Proper Motion - relates M.C. at standard epoch T_o and M.C. at epoch T,

b) Nutation - relates M.C. and T.C., both at epoch T,

c) Annual Aberration and Parallax - relates T.C. and A.P. both at epoch T,

d) Diurnal Aberration, Geocentric Parallax and Refraction - relates A.P. and O.P. at epoch T.

5.3 DUALITY PARADOX IN THE APPARENT AND OBSERVED CELESTIAL SYSTEMS

The reason there are apparent and observed systems is because the observer is not at the centre of the celestial sphere (the centre of the sun) and this must somehow be accounted for. There are two ways of making this correction, and this difference has not been made explicit earlier in these notes.

First we can retain the true celestial system (with heliocentric origin) as our coordinate system, and apply corrections to the positions of the stars. This is the approach described in sections 3.5.4 and 3.5.5 where the aberration and parallax corrections are applied to the right ascension and declination, and do not change the coordinate system. Therefore, we then say that the stars have "apparent places" or "observed places" in the true celestial system.

The second approach is to actually move the origin of the true celestial system from the centre of the sun to the centre of the earth (for the apparent system) and to the observer's position (for the observed system). This is what we have done when we related the average terrestrial to the celestial system in section 3.6 and the orbital system to the celestial system in section 4.3. In this case we called the shifted true celestial system the "apparent celestial system". In other words we have adopted the convention that

a) "true" means heliocentric,

b) "apparent" means either geocentric or corrected for the heliocentric-geocentric shift,

c) "observed" means either topocentric or corrected for the heliocentric-topocentric shift.

Two connecting parameters, Proper Motion and Refraction, do not fit in to this second scheme. Proper Motion is the changes in the positions of stars, and is different for each star. Therefore a different coordinate system would have to be defined for each star, which is nonsense. The magnitude of the refraction correction depends on the incident angle and the ambient conditions. Therefore specifying that the coordinate system follow refraction would mean a different coordinate system for each incident angle, all of which would be jumping around with the temperature and wind.

5.4 THE CONNECTIONS BETWEEN TERRESTRIAL, CELESTIAL AND ORBITAL SYSTEMS

The average terrestrial and apparent celestial systems are connected by GAST (Greenwich Apparent Sidereal Time). Note that the use of "apparent" in GAST is consistent with the convention we adopted in the previous section. That is, "apparent" means "geocentric".

The orbital and apparent celestial systems are connected by the Euler angles

a) ω = argument of perigee,

b) i = orbital inclination,

c) Ω = right ascension of the ascending node.

To summarize the differences between terrestrial, celestial and orbital systems:

a) terrestrial systems rotate and revolve with the earth,

b) celestial systems do not revolve with the earth,

c) orbital systems do not rotate with the earth.

REFERENCES

Bomford, G. (1971). Geodesy, Third Edition, Oxford University Press, London.

Brouwer, D. and G.M. Clemence (1961). Methods of Celestial Mechanics. Academic Press, New York.

Goldstein, H. (1950). Classical Mechanics. Addison-Wesley Publishing Company, Reading, Massachusetts.

Heiskanen, W.A. and H. Moritz (1967). Physical Geodesy. W.H. Freeman and Co., San Francisco.

Kaula, William M. (1966). Theory of Satellite Geodesy. Blaisdell Publishing Company, Toronto.

Krakiwsky, E.J. (1968). "Sequential Least Squares Adjustment of Satellite Triangulation and Trilateration in Combination with Terrestrial Data." Reports of the Department of Geodetic Science No. 114, The Ohio State University.

Lambeck, K. (1971). "The Relation of Some Geodetic Datums to a Global Geocentric Reference System", Bulletin Geodesique, No. 99, page 37, March.

Mueller, Ivan I. (1969). Spherical and Practical Astronomy as Applied to Geodesy. Frederick Ungar Publishing Co., New York.

Nautical Almanac Offices of the United Kingdom and the United States of America (1961). Explanatory Supplement to the Astronomical Ephemeris and Nautical Almanac. H.M. Stationary Office, London.

Schmid, H.H. (1970). "A World Survey Control System and Its Implications for National Control Networks." In Papers of the 63rd. Annual Meeting of The Canadian Institute of Surveying, Halifax, 1970.

Thompson, E.H. (1969). An Introduction to the Algebra of the Matrices With Some Applications. The University of Toronto Press, Toronto.

Veis, G. (1960). "Geodetic Uses of Artificial Satellites." Smithsonian Contributions to Astrophysics, 3 , 9. Smithsonian Institution, Washington, D.C.

Veis G. (1966). "Geodetic Interpretation" in Geodetic Parameters for a 1966 Smithsonian Institution Standard Earth, Volume 1, Edited by C.A. Lundquist and G. Veis, Smithsonian Astrophysical Observatory Special Report 200.

Wells, David E. (1971). "Matrices." Department of Surveying Engineering Lecture Notes No. 15, The University of New Brunswick, Fredericton.

Wells, D.E. and E.J. Krakiwsky (1971). "The Method of Least Squares." Department of Surveying Engineering Lecture Notes No. 18, University of New Brunswick, Fredericton.

Korn, G. A. and T. M. Korn (1968). "Mathematical Handbook for Scientists and Engineers" Second Edition, McGraw-Hill, Toronto.

Paul, M. K. (1973). "A Note on the Computation of Geodetic (Cartesian) Coordinates". Bulletin Geodesique, No. 108, page 135, June.

Hotine, M. (1969). "Mathematical Geodesy", ESSA Monograph 2, Washington, D.C.

Vanicek, P. (1972). "Physical Geodesy II", Department of Surveying Engineering Lecture Notes No. 24, University of New Brunswick, Fredericton

Mitchell, H.C. (1948). "Definitions of terms used in geodetic and other Surveys". U.S. Coast and Geodetic Survey Special Publication 242.

Fischer, I., M. Slutsky, R. Shirley, P. Wyatt (1967). "Geoid Charts of North and Central America". U.S. Army Map Service Tech. Rep. No. 62.

Smith, P.A. (1971). "North American Datum" National Academy of Sciences-National Academy of Engineering, ESSA E53(N), Washington, D.C.

Merry, C.M. and P. Vanicek (1973). "The Geoid and the Positioning of Geodetic Datums". The Canadian Surveyor, 27, No. 5 (December)

Krakiwsky, E.J., D.E. Wells and D.B. Thomson (1973). "Geodetic Control From Doppler Satellite Observations for lines under 200 km" The Canadian Surveyor, 27, 141-148.

APPENDIX A

SUMMARY OF REFLECTION AND ROTATION MATRICES

A.1 Orthogonal Transformations

The matrix equation

$$Y = A \cdot X$$

where A is a matrix and X and Y are column vectors, can be regarded as a linear transformation, in which case the matrix A is called the transformation matrix. If the two vectors X and Y have the same length, then both the transformation and the matrix are said to be orthogonal. Orthogonal matrices have the property that the product of the matrix and its transpose (or vice versa) is the identity matrix, that is

$$A^T A = A A^T = I.$$

From this property it follows that the determinant of an orthogonal matrix is either +1 or -1. There are two kinds of orthogonal transformations called reflections and rotations. The determinant of reflection matrices is -1, and the determinant of rotation matrices is +1.

There are two interpretations of the linear transformation above. The first is that the transformation describes the relationship between two coordinate systems, in which case X and Y are the same vector, but their elements refer to the two different systems. The second is that the transformation describes the relationship between different vectors X and Y in the same coordinate system. In these notes, we are interested only in the first interpretation.

A.2 Right and Left Handed Cartesian Coordinate Systems

A three dimensional Cartesian coordinate system can be orthogonally transformed in only six different ways. It can be rotated about each of its axes. Each of its axes can be reflected. In such a coordinate system, the vectors X and Y will have only three elements. Let us define the axis to which the first, second, and third elements of X and Y are referred as the 1-axis, 2-axis, and 3-axis respectively (we could equally well label them the x_1, x_2, x_3 axes or x, y, z axes).

These three axes may define either a right-handed or a left-handed coordinate system. Right handed systems follow the right hand rule: if the fingers of the right hand are curled around any axis so that the thumb points in the positive direction, then the fingers will point from a second axis to the third axis, numbered in cyclic fashion. Grasping the 1-axis, the fingers point from the 2-axis to the 3-axis. Grasping the 2-axis, the fingers point from the 3-axis to the 1-axis. Grasping the 3-axis, the fingers point from the 1-axis to the 2-axis. Left-handed coordinate systems follow the left hand rule, which differs from the above only in that the left hand is used.

A.3 Reflections

If we denote a reflection of the kth axis by P_k, then the following expressions define the three reflection matrices:

$$P_1 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Note that reflection matrices commute (e.g. $P_2P_3 = P_3P_2$), so that it makes no difference in what order a sequence of reflections are performed. Note also that an odd number of reflections changes the handedness of the coordinate system.

A.4 Rotations

If we denote a rotation of angle θ about the k^{th} axis by $R_k(\theta)$, then the following expressions define the three rotation matrices:

$$R_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

$$R_2(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_3(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that rotation matrices do not commute. The product of several rotations is performed from right to left, for example in

$$R_1(\alpha) \quad R_2(\beta) \quad R_3(\gamma)$$

the rotations are performed about the 3-axis of the original system, the 2-axis of the transformed system, and the 1-axis of the doubly transformed system, to yield the final triply transformed system.

If the rotation angles are all so small that their cosines can be assumed to be unity, then the rotation matrices become commutative. This is the case for differential rotations, for example.

The above expressions define positive rotations, which are right-hand rotations for right-handed coordinate systems and left-hand rotations for left-handed coordinate systems. A right-hand rotation is related to the right hand rule given above: if the fingers of the right hand are curled around the rotation axis so that the thumb points in the positive direction, then the fingers curl in the direction of a right hand rotation. A similar statement for left hand rotations is obvious.

A.5 Inverse Transformations

The inverse of a transformation A (denoted A^{-1}) is the transformation which returns conditions to their original state, that is

$$A^{-1} A = A A^{-1} = I.$$

Reflections are self-inverse, that is

$$P_k^{-1} = P_k$$

$$P_k P_k = I$$

Common sense tells us that the inverse of a positive rotation is a negative rotation, that is

$$R_k^{-1}(\theta) = R_k(-\theta)$$

and this conclusion is verified by taking the orthogonal property

$$A^T A = I$$

from which it is evident that for orthogonal matrices

$$A^{-1} = A^T$$

and for each of the above expressions for rotation matrices it can be shown that

$$R_k^T(\theta) = R_k(-\theta).$$

Applying the rule for the inverse of products

$$[A B]^{-1} = B^{-1} A^{-1}$$

we have

$$[R_j(\alpha) R_k(\beta)]^{-1} = R_k^T(\beta) R_j^T(\alpha) = R_k(-\beta) R_j(-\alpha)$$

A product transformation consisting of one rotation and one reflection commutes only if the rotation and reflection refer to the same axis, that is

$$P_j R_k = R_k P_j \quad \text{if } j = k$$

otherwise

$$P_j R_k = R_k^{-1} P_j \quad \text{if } j \neq k.$$