AN INTRODUCTION TO GEODESY

ART STOLZ

MONOGRAPH 16

SCHOOL OF
GEOMATIC ENGINEERING

An Introduction to

Geodesy

Art Stolz

Second Edition

January 2001

(Reprint May 2009)

Contents

Preface

The practice of geodesy has changed dramatically in the last decade mainly through the impact of satellite methods. For example, positioning for most geodetic applications is now almost exclusively performed with GPS and the finer structure of the earth's gravity field is increasingly being determined from satellite altimetry. These changes are reflected in the teaching of geodesy in the School of Geomatic Engineering (formerly the School of Surveying) of The University of New South Wales.

This monograph is based on a course called "Introduction to Geodetic Science" which has been taught at The University of New South Wales for a number of years. The course is designed for First Year students who, it is assumed, know nothing about the subject but who have a basic grounding in calculus, the manipulation of matrices and physics. The course is taught in fourteen two-hour lectures and about ten one-hour tutorials.

Chapter One introduces geodesy by describing what modern geodesy is, tracing the development of geodesy through time, relating geodesy to other sciences, giving some of the current applications of geodesy, briefly reviewing geodetic activities in Australia and providing a selection of international and local journals concerned with geodetic problems.

Chapter Two deals with the earth's gravity field and its variations over the earth's surface, in particular those variations related to the shape of the sealevel surface or geoid. It is shown how the geoid may be determined from terrestrial gravity data and a gravity model for the earth using Stokes' theorem and how the complex shape of the geoid may be built up from spherical harmonic coefficients obtained from satellite methods.

Chapter Three is devoted to time and time keeping. The principles of atomic clocks are described. The different time scales that are used in geodesy, and which also occur in astronomy, surveying and navigation, are discussed. Finally, the relationship between these time scales is given.

Chapter Four is concerned with the motion of a satellite about a parent body. Kepler's laws of planetary motion, which to a first order also apply to artificial earth-satellites, are given. The elements which are commonly used to

describe the geometry of an orbit and to relate the orbit to a space-fixed reference frame are defined. The various perturbing effects which cause an orbit to depart from the ideal described by Kepler's laws are then discussed. It is shown how some of these perturbations may be used to determine the earth's shape.

Chapter Five discusses the coordinate types which occur in geodesy. The two most important geodetic reference coordinate systems, the geocentric-equatorial inertial or space-fixed system and the geocentric-equatorial rotating or earth-fixed system, are described and it is shown how coordinates expressed in the earth-fixed system may be transformed to coordinates in the space-fixed system, and vice versa.

Chapter Six deals with the terrestrial geodetic positioning methods and the measurement of terrestrial gravity. Horizontal and vertical control surveys are discussed. It is shown that levelled heights are path dependent. The concepts of orthometric, dynamic and normal heights are introduced to overcome this problem. Finally, the relationship between the mean sealevel surface and the geoid is explained.

Chapter Seven introduces the space geodetic methods currently used for precise positioning and earth gravity field determination. The principles of satellite laser ranging, GPS, VLBI and satellite radar altimetry are explained. It is shown how these methods are used to determine geodetic parameters. Measurement errors are discussed and methods of removing them are presented.

This monograph teaches the rudiments of geodetic positioning and, to a lesser extent, earth gravity field determination as performed today. As such, the emphasis is on satellite methods. Numerous numerical examples are given which the reader can work through to gain a better grasp on the subject.

The finishing touches to the first edition of this monograph were applied whilst the author held a Visiting Fellowship at the Research School of Earth Sciences of the Australian National University where Clementine Krayshek drew the figures.

In this second edition numerous revision exercises are given at the end of each chapter. Material on geodetic datums, transformation between datums and how to obtain orthometric height from space geodetic methods has been added. Also, more worked examples appear throughout, new reference material is included and typographical and other errors have been corrected. Most importantly, however, we provide an index.

Art Stolz
July 1996

1 Introduction

1.1 WHAT IS GEODESY?

A dictionary might define geodesy, in modern usage, as being *that branch of mathematics which determines the figures and areas of large portions of the earth's surface, and the figure of the earth as a whole.* This definition, which suggests that geodesy is no more than the study of the geometry of the earth, does not, however, indicate the full scope or current extent of the science of geodesy. There are at least two further essential components of geodesy: (1) the study of the earth's gravity field and its temporal variation, and (2) the study of the earth's rotational motion. Both are essential in the practice of geometrical geodesy since one of the fundamental reference surfaces is the *geoid*, or the equipotential surface corresponding to mean sealevel, and since positions on the earth's surface are usually determined from observations made with respect to a reference system fixed to the stars.

As a rough generalization, it can be said that there are three principal types of activity in geodesy: *terrestrial* or *classical geodesy* based on measurements of the geometry and gravity of points at the surface of the earth; *space geodesy* based on observations of and from artificial satellites; and *theoretical geodesy* which is concerned with the analysis and interpretation of the measurements. Space geodesy also covers the use of radio interferometry to obtain geometrical information from observations of extraterrestrial radio sources.
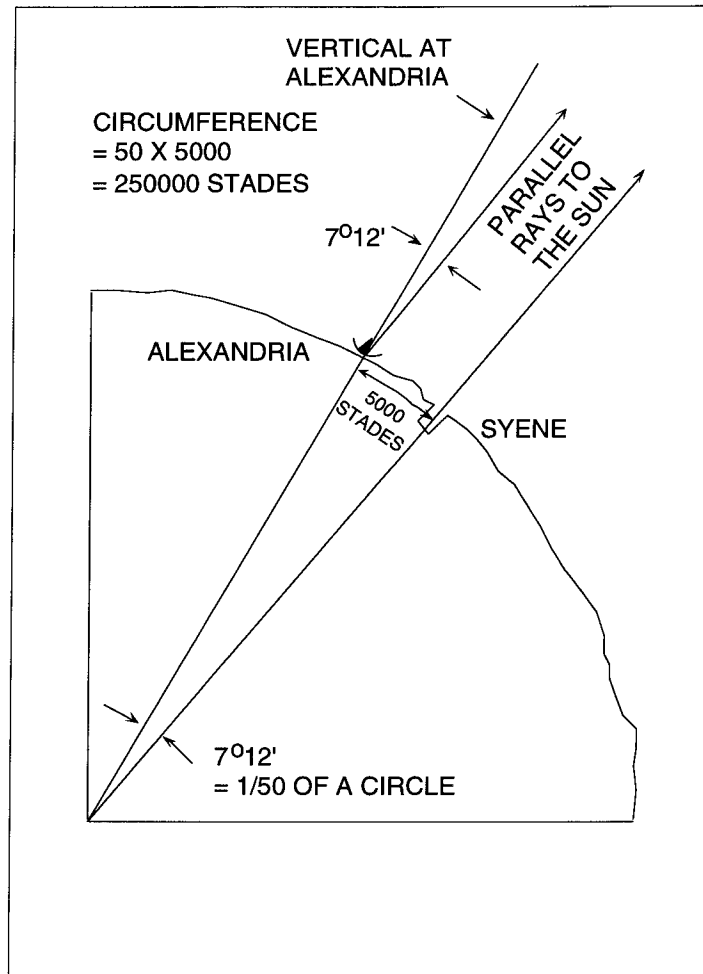
Figure 1.2-1
Erastothenes' experiment.

Today, most practical geodesy is performed using space geodetic methods. One strives to attain the highest accuracy possible. Relative positioning accuracies of a few parts in 10^8, which is equivalent to a few millimetres over a 100 km long line, have been achieved with satellite laser ranging and radio interferometric methods. Relative accuracies 1 part in 10^7 (1 cm for a 100 km long line) are now almost routinely achievable with the satellites of the Global Positioning System (GPS).

1.2   HISTORICAL DEVELOPMENT OF GEODESY

The history of the science of geodesy begins with the idea of a spherical earth. As far as is known, the greek scientist Pythagoras was one of the first to propose a spherical shape for the earth in the sixth century BC. Aristotle took up this idea and gave it an observational basis by noting the apparent movement of the
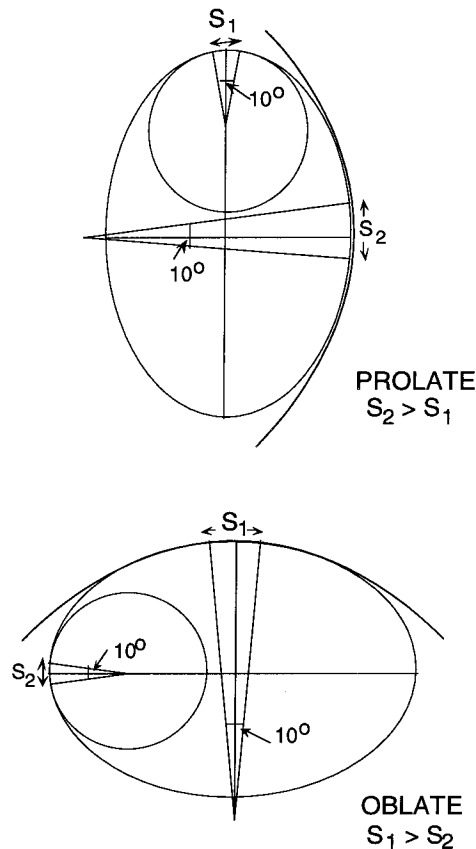
Figure 1.2-2
Oblate and prolate ellipsoid of revolution.

stars, the circular shadow of the earth during a lunar eclipse and the depression of the horizon with increasing distance. However, it took about three centuries before a serious attempt was made to measure the radius of the earth. Erastothenes, the director of the famous library at Alexandria, determined the earth's radius by observing the elevation of the sun at noon on the day of the summer solstice in Alexandria. Since at that time the sun was at its zenith position in Aswan, and since he was aware of the relative position of the two cities and that they were located on approximately the same meridian, Erastothenes was able to calculate the radius of the earth from simple geometry (Figure 1.2-1). Remarkably, considering his poor measuring tools, his solution was only 16% in error.

With the decline of the greek empire and the spread of Christianity in Europe, scientific study waned and it was not before the end of the Middle Ages, that the voyages by da Gama and Columbus revived the interest for determining the shape of the earth. The idea that the earth was flat was finally buried with their discoveries and new attempts were made to determine the earth's radius. The frenchman Fernel was the first to provide a new estimate in 1525. Fernel observed the elevation of the sun in Paris and Amiens. Using astronomical tables and the distance between the two cities, measured by an odometer, he obtained a value which was only out by about 1%. The development of new

instruments made other and more accurate techniques possible. The most important of these for geodesy was the theodolite. Snell, a professor of mathematics in Leiden, used it in 1615 to measure by triangulation the distance between the dutch cities of Alkmaar and Bergen op Zoom. The scale of the network was determined from a baseline observed with a surveyor's chain. With astronomic latitude observations at the extremeties of the triangulation chain, the radius was determined with an error of roughly 3%.

The discovery of his laws of motion, led Newton to conclude that gravity, as observed by a pendulum, must decrease in magnitude from the poles towards the equator, as a result of the centrifugal force of the earth's rotation. Newton, or Picard, also hypothesized that the earth was an oblate ellipsoid, instead of a perfect sphere. To test this hypothesis, the french Academy of Science commissioned Cassini to perform a triangulation from Dunkerque, in the north, to the Pyrenees, in the south of France. The splitting up of the meridian arc in several sections would show whether the length of a degree depended on latitude, as was the case with an ellipsoid. Surprisingly, Cassini came to a conclusion opposite to Newton, that is, the earth was a prolate ellipsoid, flattened at the equator rather than the poles (Figure 1.2-2). To resolve the matter, the french Academy sent out two expeditions in 1736, one to Lapland and the other to Peru, to measure the length of a degree of latitude in two distinctly different places. From these expeditions and from the many others which followed, Newton's hypothesis that the earth was an oblate ellipsoid, was confirmed.

An important and fundamental discovery was made during the Peru expedition. Bouguer noticed variations in gravity that could not be accounted for by changes in elevation or latitude. These regional gravity variations provided the first evidence for a nonuniform density distribution within the earth. Also, in 1738, Clairaut published the mathematical relationship between the gravity flattening and geometrical flattening of the earth. This implied, that the earth's geometrical flattening could be determined from gravity measurements. Stokes, in 1849, obtained a more generalized expression relating gravity measurements made at sealevel to the departure of the earth's shape from a sphere.

Laplace, Gauss, Bessel and others soon recognized that at sufficiently high observational accuracy one could no longer ignore the deviation of the plumbline, to which the measurements refer, from the ellipsoidal normal. This led to the concept of the geoid and efforts for its determination. The broadscale features of the geoid are now well known from satellite methods (Figure 1.2-3). However, the determination of its fine structure remains an important problem of geodesy, particularly with the present exacting requirements of oceanography. Nevertheless, the oblate ellipsoid continues to serve as the basic reference surface for geodetic computations.
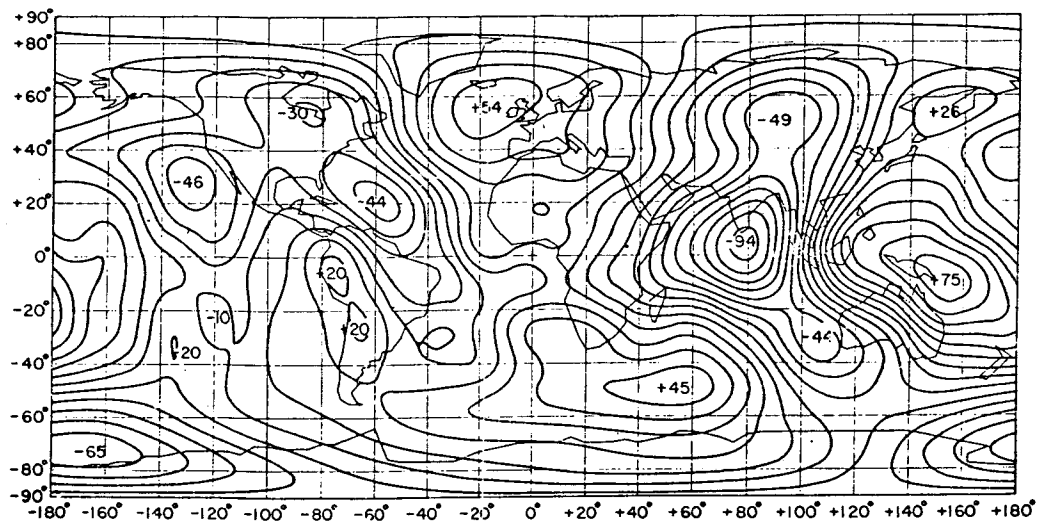
Figure 1.2-3
Satellite geoid.

1.3 GEODESY AND OTHER SCIENCES

The practice of geodesy necessarily includes several topics that are also studied in connection with other fields of science and technology. The following examples illustrate the wide scope of geodesy and the diversity of interests of geodesists:

- Earth and ocean tides affect both the shape of the ocean surface and the gravity field. Their measurement and interpretation may therefore be regarded as essential to geodesy as well as to geophysics and oceanography.

- Refraction of electromagnetic waves in the earth's atmosphere affects geodetic measurements that are made at various wavelengths of the electromagnetic spectrum, and so the interests of geodesists overlap with those of the meteorologist, astronomer and ionospheric physicist.

- The study of the gravity field by surface measurements and by satellite methods is of direct relevance to geology since it reveals the existence of otherwise inaccessible structures at various depths inside the earth. An accurate knowledge of the geoid is also required in satellite altimetry to allow the identification of the undulations of the ocean surface caused by currents and winds.

- Local, regional and global movements of the earth's crust result in changes in the positions of geodetic reference points, and so geodesy contributes to the study of crustal dynamics, a term that is used to study the relationship between earthquakes and plate tectonic motion. The space geodetic techniques of satellite laser ranging, radio interferometry and GPS, are providing direct determinations of the current rates of motion of the earth's crustal plates.

- The rotation of the earth around its axis no longer provides the standard of time, but a knowledge of the orientation of the earth with respect to a space-fixed reference frame is necessary for interpreting many types of astronomical observations. Thus, the study of the rotation of the earth is of fundamental importance in astronomy. Moreover, variations in the rotation rate and the orientation of the rotation axis provide valuable data on the interior of the earth and about the interactions between the crust, oceans and atmosphere. These studies, apart from providing useful data for geophysics, oceanography and meteorology, are of significance to theories of the origin and evolution of the earth-moon system. They may also provide a measure of the possible variation in the constant of gravitation.

1.4 APPLICATIONS OF GEODESY

Apart from its applications to other sciences, geodesy is of some importance in modern life. The most direct applications are to surveying and mapping, but there are also indirect applications that help justify the substantial costs of the basic geodetic measurements and associated data processing.

The techniques of surveying allow determination of precise relative positions of points over limited regions, but coordinates assigned to points near the edge of one such region are usually found to differ significantly from those given by another independent survey of a neighbouring region in those areas where the two regions overlap, and both sets of coordinates must be transformed to a common reference surface so as to eliminate ambiguities. Such differences can be of considerable commercial significance, for example, in oilfields that cross boundaries which have been defined in terms of these coordinates.

The satellite techniques for accurate postioning and navigation depend for their success on information that has been obtained as a result of past geodetic investigations and on current determinations of the rotation of the earth. In turn, these techniques are useful in geodesy since they provide the most economical way of establishing a network of points, the positions of which are known in the adopted terrestrial reference coordinate system.

It is possible that geodetic measurements in regions prone to earthquakes will provide information that may indicate where and when major earthquakes will occur; for instance, long lines straddling the San Andreas fault in California are

being repeatedly measured by space techniques to determine where slipping is occurring and where stress is building up.

Accurate positioning and other geodetic information are essential in many engineering projects, such as the relocation of drilling rigs at sea and the construction of long tunnels and canals. Furthermore, the observational and computational methods developed for geodetic networks are being applied to monitor motions in and around large civil engineering structures, such as dams. These techniques are also being used in the very precise setting out and monitoring of large scientific instruments, such as particle accelerators.

Information about geological sructures obtained from gravimetry is used in the location of oil, mineral and ore deposits. The possibility of detecting storm surges at an early stage, by using satellite altimetry with an accurate orbit and geoid, could provide warning of the risk of flooding of coastal areas.

Correct interpretation of the images obtained by earth resources satellites such as Landsat and Spot depend on determining the proper relationship between the images and the actual surface of the earth which requires the prior establishment of a network of control points whose positions are known with respect to a standard reference coordinate system.

The regular monitoring of the rotation of the earth is used to derive a scale of universal time which is used for navigation by ships and aircraft around the globe.

1.5 GEODESY IN AUSTRALIA

Research in the science and application of geodesy in Australia is carried out in a variety of institutions as a subsidiary activity rather than a principal activity; there is no national geodetic agency or geodetic institute to play a central role, as is the case in many other countries. Funding of geodetic activities is spread between several federal and state government departments, the Universities and the Australian Research Council. Similarly, there is no major learned or professional society that explicitly provides for the presentation and publication of geodetic research.

The Australian Land Information Group (AUSLIG) of the Department of Administrative Services in Canberra is reponsible for maintaining the first-order triangulation and levelling networks in the ACT, and for geodetic activities in the Australian Antarctic Territory and certain offshore islands, such as Norfolk Island. AUSLIG operates a space geodetic observatory in the Orroral Valley near Canberra which, amongst other things, tracks near-earth satellites and retroreflectors on the moon, and disseminates time in Australia. Each state government is responsible for maintaining its own geodetic networks. Coordination of this activity is maintained by an intergovernmental committee.

Other government organizations with a direct interest in geodesy are the Directorate of Survey-Army and the Australian Navy hydrography unit.

Geodesy is taught as a postgraduate activity at several universities in Australia, principally the University of South Australia and The University of New South Wales. The most active research group is in the School of Geomatic Engineering, the University of New South Wales. Research activites there include or have included the determination of large-scale crustal motion across the Java Trench and in New Guinea, computing the fine structure of the geoid in the Australian region, sealevel monitoring with GPS and, precision altimetric measurements of the earth's land surface. The CSIRO Division of Radiophysics performs geodetic radio interferometric experiments in collaboration with the NASA Jet Propulsion Laboratory and geodetic agencies in other countries, though not on a regular basis.

The Australian Academy of Sciences provides a valuable, but informal, forum in support of geodesy in Australia as a byproduct of its responsibility for providing a committee to act as a link to the International Association of Geodesy. Geodesists drawn from the various organizations with an interest in geodesy meet in the Geodesy Subcommittee of the National Committee for Solid-Earth Sciences, to discuss Australian participation in international cooperative projects and geodetic research.

1.6   GEODETIC LITERATURE

Useful reference books on geodesy are given at the end of each chapter of this monograph. Among the international technical journals *Bulletin Géodesiqué* and *Manuscripta Geodaetica* are concerned exclusively with geodetic problems. In 1995 these two journals were discontinued and merged into a single publication called the *Journal of Geodesy*. State-of-the-art articles and reviews appear in , for example, *Reviews of Geophysics and Space Physics*, *Journal of Geophysical Research* and *Geophysical Research Letters* all of which are published by the American Geophysical Union in Washington, DC. *Geomatics Research Australasia* (formerly the *Australian Journal of Geodesy, Photogrammetry and Surveying*) publishes results of geodetic research mainly for the Australasian region.

REFERENCES

FISHER, I., 1975. The figure of the earth - changes in concept, *Geophysical Surveys*, **2**, 3-54.

INTERNATIONAL ASSOCIATION OF GEODESY, 1992. The Geodesist's Handbook, *Bulletin Géodesiqué*, **66** (2), 260pp.

SMITH, J.R., 1988. *Basic Geodesy*, Landmark Enterprises, Rancho Cordova, CA, 151pp.

TORGE, W., 1991. *Geodesy* (2nd edition), Walter de Gruyter, Berlin, 264pp.

UNITED STATES NATIONAL RESEARCH COUNCIL, 1978. *Geodesy: Trends and Prospects*, Committee on Geodesy, National Academy of Sciences, Washington, DC, 86pp.

WILKINS, G.A., 1983. The scope and value of geodesy with notes on geodetic activities in the United Kingdom, *Quarterly Journal of the Royal Astronomical Society*, **24**, 106-112.

VANICEK, P. & KRAKIWSKY, E.J., 1982. *Geodesy: The Concepts*, North-Holland Publishing Company, Amsterdam, 691pp.

EXERCISES

1.1   What is geodesy? Why do we need to study it?

1.2   Geodesy is a global discipline. Explain why?

1.3   What is the relationship between geodesy and surveying?

1.4   Explain the requirement for high accuracy in geodesy.

1.6   The oblate ellipsoid continues to serve as the reference surface for geodetic computations. Why?

1.7   Geodesy serves many other disciplines. Discuss this statement.

1.8   Give the main goals of geodesy.

1.9   You have been asked to give a short lecture on geodesy to senior high school students who have never heard of it. Your main aim is to get them interested in the subject. What would you tell them?

2 The Earth's Gravity Field

2.1 INTRODUCTION

One of the most familiar facts about the earth is that a body released near it will fall with increasing velocity. The rate of increase of velocity is called the *acceleration of gravity*, **g**, and is the same for all bodies at a given point on the earth. Newton formulated the principle of universal gravitation by deduction from Kepler's laws of planetary motion, showing that these laws were evidence of a force between each planet and the sun.

In the case of a body on the earth, the force of attraction is determined by the product of the earth's mass and the mass of the body, and the distance between the body and the earth's centre. If the earth were a uniform, nonrotating sphere, the force on a body at a given distance would be everywhere the same, and there would be a single constant value for **g**. However, the earth is nonuniform, nonspherical and rotating, and all these facts contribute to the variations in **g** over its surface.

Virtually all geodetic measurements are influenced by the force of gravity and this must be taken into account when calculating geodetic parameters. Moreover, those variations in gravity which are related to the departure of the earth from a spherical form are of particular interest in geodesy. Measurements and analyses of the variation in gravity also form a powerful branch of geophysics; the determination of the structure of the earth's interior. Variations which reflect the nonuniform density of the earth can be used to infer the
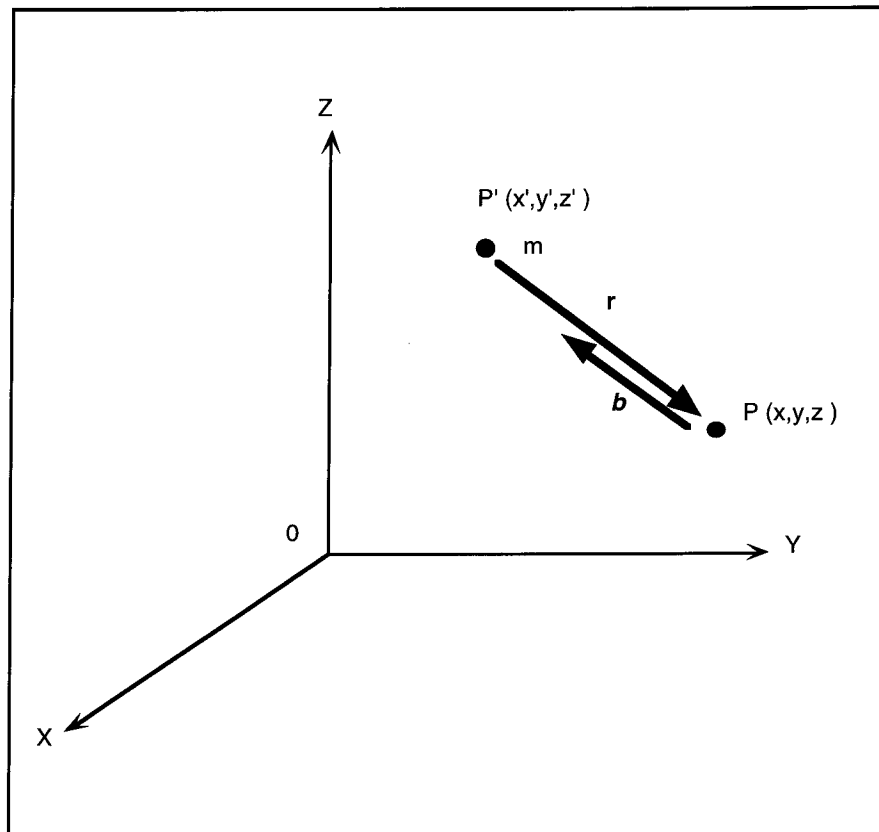
Figure 2.3-1
Newton's law of gravity.

pressure structure beneath the surface, and these are of interest to the geologist.

2.2 UNITS

In the SI system the unit of gravity is the ms^{-2}. The magnitude of **g** varies from 9.78 ms^{-2} at the equator to 9.83 ms^{-2} at the poles. For investigations of the earth's shape, or its internal structure it is necessary to measure variations of 10^{-5} ms^{-2} or less and it is convenient to introduce the units:

$$1 \; \mu\text{ms}^{-2} = 10^{-6} \text{ ms}^{-2} \quad \text{and} \quad 1 \text{ nms}^{-2} = 10^{-9} \text{ ms}^{-2}.$$

In geodesy and geophysics, we often find the auxillary units:

$$1 \text{ mGal} = 10^{-5} \text{ ms}^{-2} \quad \text{and} \quad 1 \; \mu\text{Gal} = 10^{-8} \text{ ms}^{-2}.$$

These are derived from the unit Gal (1 Gal = 1 cms^{-2}), named after Galileo. In official and commercial communications of many countries, these auxillary units are no longer permitted. However, in scientific circles they are still in partial use as is the μms^{-2}.

2.3 FUNDAMENTAL CONCEPTS

Newton's Law of Gravitation

Newton's law of gravitation for two particles of mass, m_1 and m_2, separated by distance, r, is:

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

(2.3-1)

where G = (6.67259 ± 0.00085)x10^{-11} m^{3}kg^{-1}s^{-2}, is the gravitational constant and, F, is the force on either mass, and is directed along the line joining the masses. Hence, a unit mass located at the attracted point P (Figure 2.3-1) in the gravitational field experiences a gravitational acceleration, b, due to the mass element, m, at the attracting point P', ie.:

$$b = -G \frac{m}{r^2} \frac{r}{r}$$

(2.3-2)

where b lies on the line joining P and P' and is directed towards P'.

Potential

A body such as the earth, composed of an infinite number of mass elements, produces a gravitational acceleration on the unit mass, situated at the attracted point, which is computed by summing the individual accelerations vectorially. The computations are simplified, if we change from a vector field to a scalar field. Now, a mass in the presence of an attracting body of mass, m, has energy by virtue of the attraction. This energy:

$$V = - \frac{Gm}{r}$$

(2.3-3)

known as *potential energy*, can be evaluated by considering the mass to have been brought from infinity and calculating the work done on it in the process. We note that in geodesy we customarily make V positive. The potential has the dimension of work/unit mass and the units are m^{2}s^{-2}. It is a relatively simple
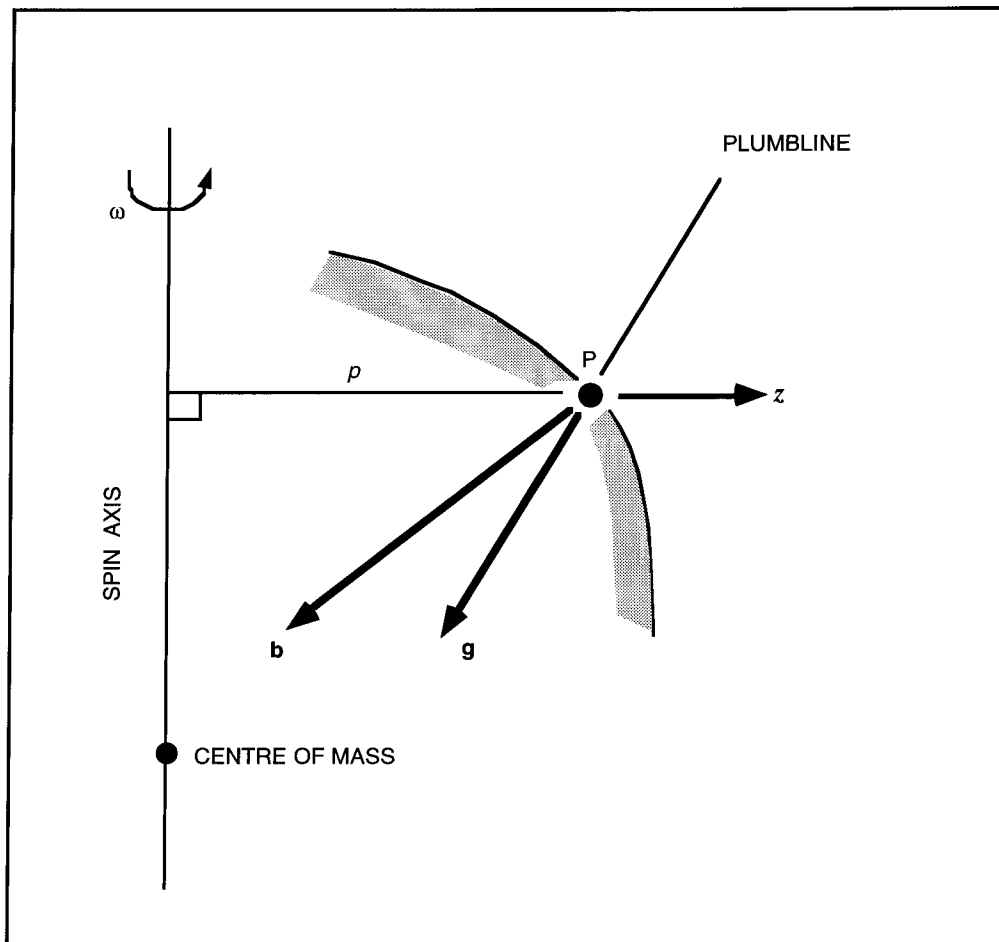
Figure 2.3-2
Centrifugal force due to the earth's rotation.

matter to show that the force of attraction, **F**, can be obtained from the potential energy by differentiation ie.:

$$\mathbf{F} = \text{grad } V = (\frac{\partial V}{\partial X}, \frac{\partial V}{\partial Y}, \frac{\partial V}{\partial Z})$$ (2.3-4)

Equipotential Surfaces

A gravitational field can be represented by surfaces over which the potential, V, is constant. These are known as *equipotential surfaces*. The force vectors are everywhere normal to these surfaces, so that there is no component of force along them. Thus the surface of a liquid in the gravity field coincides with an equipotential surface and for this reason the potential and equipotential surfaces are of great importance in the study of the sealevel surface of the earth or geoid.

Centrifugal Acceleration

The force acting on a body at rest on the earth's surface is the resultant of the gravitational force and the centrifugal force of the earth's rotation. Here we assume that the angular velocity, ω, of the earth is constant and that the rotation or spin axis is fixed with respect to the earth. The centrifugal acceleration on unit mass is:

$$z = \omega^2 p \tag{2.3-5}$$

This acts in an outward direction perpendicular to the spin axis (Figure 2.3-2). The earth's angular velocity, determined from astronomical observations, is:

$$\omega = \frac{2\pi}{86\ 164.10\ s} = 7.292\ 115 \times 10^{-5}\ rad.\ s^{-1}.$$

If the Z-axis of an earth-fixed X,Y,Z-system coincides with the earth's spin axis, then:

$$\mathbf{p} = \begin{pmatrix} X \\ Y \\ 0 \end{pmatrix}, \qquad p = |\mathbf{p}| = \sqrt{X^2 + Y^2}$$

The *centrifugal potential*, Φ, is:

$$\Phi = \Phi(\mathbf{p}) = \frac{\omega^2}{2} p^2 \tag{2.3-6}$$

with:

$$z = grad\ \Phi \tag{2.3-7}$$

For points on the equator, the centrifugal potential, $\Phi = 1.1 \times 10^5\ m^2 s^{-2}$ and the centrifugal acceleration, $z = |z| = 0.03\ ms^2$ corresponding to $\approx 0.3\%$ of gravitational acceleration, while at the poles, $\Phi = 0$ and $z = 0$.

Gravity Acceleration

The gravity acceleration, or *gravity*, $\mathbf{g}$, is the resultant of gravitation, $\mathbf{b}$, and the centrifugal acceleration, z:
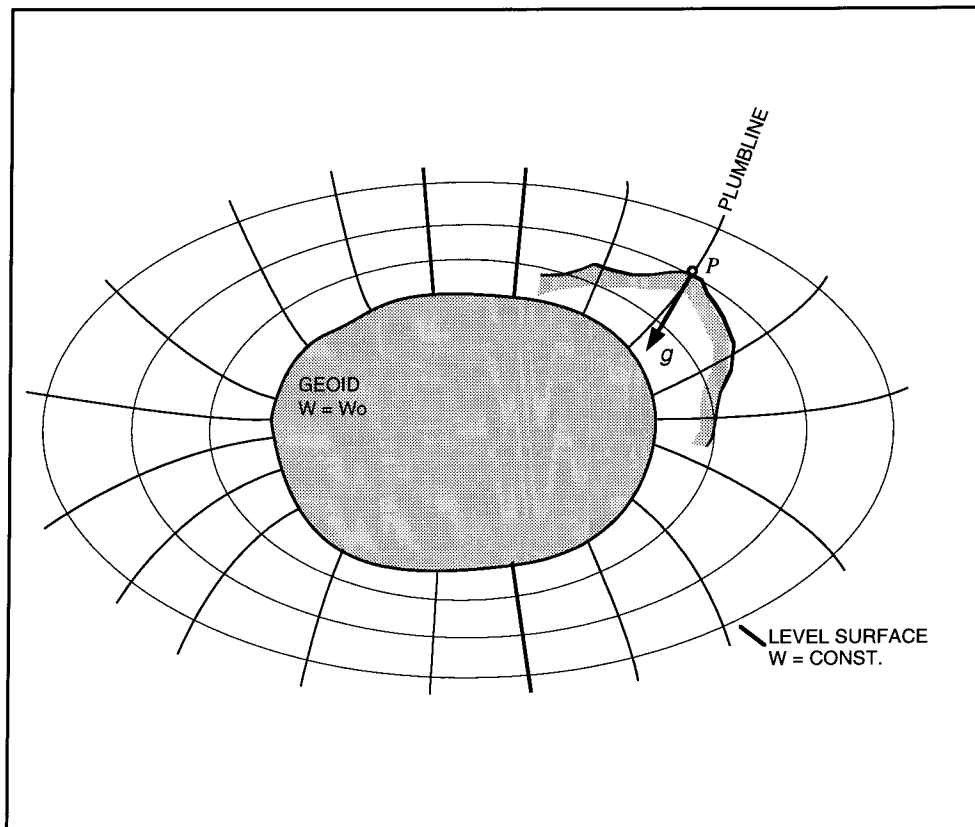
Figure 2.4-1
Convergence of level surfaces.

$$\mathbf{g} = \mathbf{b} + \mathbf{z} \tag{2.3-8}$$

The direction of **g** is known as the direction of the *plumbline* (ie. the opposite direction to the vertical) and the magnitude, $g = |\, \mathbf{g}\, |$, is called the intensity-of-gravity or, often, just gravity. The *gravity potential* becomes:

$$W = V + \Phi \tag{2.3-9}$$

and the gravity acceleration is given by:

$$\mathbf{g} = \text{grad } W \tag{2.3-10}$$

2.4 LEVEL SURFACES AND PLUMBLINES

The surfaces of constant gravity potential, W = const., are called equipotential, level or *geopotential surfaces* (geops) of gravity. The potential difference, dW, between two level surfaces differentially separated by displacement, ds, is:

$$dW = \mathbf{g} \cdot \mathbf{ds} = g \, ds \, \cos\theta \qquad (2.4\text{-}1)$$

where θ is the angle between **g** and **ds**.

We see that the derivative of the gravity potential in a particular direction equals the projection of the gravity along this direction. If we take **ds** along the level surface, then it follows from the fact that dW = 0, that **g** is perpendicular to the level surface. Accordingly, the level surfaces are intersected by the plumblines which are perpendicular to them and the tangent to the plumbline is the direction of the plumbline. If **ds** is directed along the outward pointing normal, **n**, to the level surface, then, since $\cos\theta = -1$, we have:

$$dW = g \, dn \qquad (2.4\text{-}2)$$

This equation provides the link between the potential difference, which is a physically measurable quantity, to the difference in height, which is a geometric quantity, of adjacent level surfaces. We return to this problem later on (see Section 6.3).

If g varies on a level surface, then the distance dn to a neighbouring level surface must also change. Therefore, the level surfaces are not parallel and the plumblines are not plane curves. Also, as a result of an increase on 0.05 ms^{-2} in gravity from the equator to the poles, the level surfaces of the earth converge towards the poles (Figure 2.4-1). The relative decrease of the distance between two level surfaces near the earth from the equator to the pole is on the order of 5x10^{-3}. Thus, two level surfaces which are 100.0 m apart at the equator are separated by only 99.5 m at the poles.

2.5 TEMPORAL VARIATIONS OF THE EARTH'S GRAVITY FIELD

Time dependent variations of **g** are caused by the lunar and solar gravitational forces acting on different parts of the rotating earth, in combination with the effects of the orbital motion of the moon around the earth and of the earth around the sun. These produce variations in the terrestrial gravity field on the order of 10^{-7} **g**. Other variations of the gravity field with time, caused by

terrestrial mass movements, such as the shifting mass of the atmosphere, generally are at least an order of magnitude smaller than solar and lunar gravitational effects.

2.6 GRAVITY AND THE SHAPE OF THE SEALEVEL SURFACE

The value of g varies over the earth for a number of reasons. Here, we give particular attention to those variations which are related to the earth's shape. The departure of the form of the sealevel surface from a sphere leads to variations in g, even at sealevel; conversely, a study of these variations is of great assistance in determining these departures. The study of the shape of the sealevel surface is intimately related to many of the major problems of geodesy.

We consider the waters of the oceans as freely moving homogeneous matter subjected only to the force of gravity of the earth. Upon attaining a state of equilibrium, the surface of these idealized oceans assumes a level surface of the earth's gravity field. This level surface is the *geoid*. The equation of the geoid is:

$$W = W_o = \text{const.} \tag{2.6-1}$$

We may regard the geoid as being extended under the continents by a series of narrow canals cut into the earth. Also, the land surface of the continents is defined by its height above sealevel, which is the height above the geoid. Accordingly, another way of visualizing the geoid beneath the continents is that surface everywhere at a depth equal to the measured height below the land surface. Where there are local variations in g, due to internal density anomalies, the geoid is distorted. Most of these distortions are rather limited in extent, and all of them are very small in amplitude (<100 m - see Figure 1.2-3) compared to the earth's radius.

2.7 NORMAL GRAVITY

We customarily introduce a gravity model reference system to determine the geoid. This model is called the *normal gravity field*. The source of this field is an earth model which represents a good fit to the earth's surface and to the earth's external gravity field. The earth's surface may be closely approximated by an ellipsoid of revolution with flattened poles. This ellipsoid has a simple analytical representation making it well suited as a model earth. Later on, we shall see that the shape of the ellipsoid is defined by its semi-major axis, a, and the flattening, f (see Section 5.2). The normal gravity field is obtained by introducing the mass of the earth and its angular velocity as additional parameters. Normal
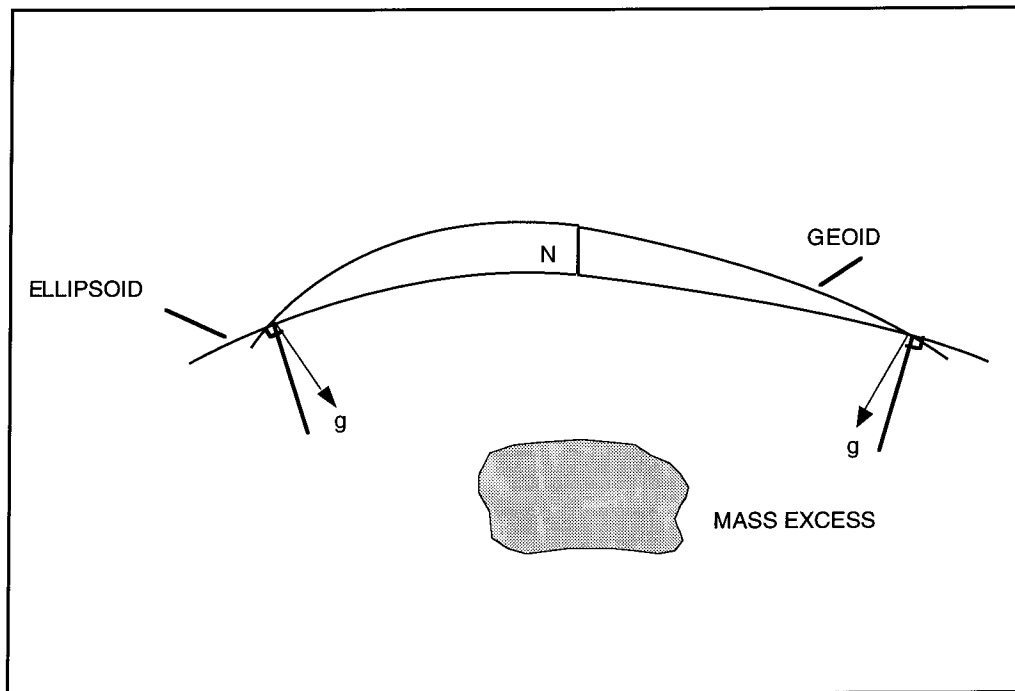
Figure 2.8-1
Geoidal undulation and deflection of the vertical.

gravity is then the combined effect of the gravitation and rotation of this ellipsoid. The gravity field in the space outside the ellipsoid is uniquely defined if we further require the ellipsoid to be a level surface of its own gravity field. This body is known as the *level* or *equipotential ellipsoid*. If the ellipsoid parameters are now assigned values which closely match those of the real earth, then we have the optimum approximation to the geometry of the geoid and to the external gravity field. This body is the *mean earth ellipsoid*.

It can be shown that gravity, γ_0, at latitude, ϕ, on the level ellipsoid is given by:

$$\gamma_0 = \gamma_e (1 + B_2 \sin^2\phi - B_4 \sin^2 2\phi) \tag{2.7-1}$$

where γ_e is normal gravity at the equator and, B_2 and B_4, are constants which are on the order of 5×10^{-3} and 10^{-6}, respectively. For specific values of B_2 and B_4 this equation becomes the *International Gravity Formula*.

Example 2.7-1

By *how much* does gravity increase in going from a point at sealevel at latitude $50°$ to one at sealevel at latitude $55°$? *Take* $\gamma_e = 9.78$ ms^{-2}, $B_2 = 5.3 \times 10^{-3}$ and $B_4 = 0$.

<u>Solution</u>

Equation (2.7-1) becomes:

$$\gamma_o = \gamma_e \ (1 + 5.3 \times 10^{-3} \sin^2\phi)$$

This gives:

$$\gamma_{50^\circ} = 9.78 \ (1 + 5.3 \times 10^{-3} \sin^2 50^\circ) = 9.810417 \ \text{ms}^{-2}$$

and:

$$\gamma_{55^\circ} = 9.78 \ (1 + 5.3 \times 10^{-3} \sin^2 55^\circ) = 9.814781 \ \text{ms}^{-2}$$

Thus:

$$\Delta g = \gamma_{55^\circ} - \gamma_{50^\circ} = 4364 \ \mu\text{ms}^{-2}$$

2.8   GEOIDAL UNDULATION

Mass anomalies, or lateral variations in density inside the earth, cause g at sealevel to differ from the values predicted by the international gravity formula. The effect of these variations on the geoid is that over a region of mass excess there is an additional or *disturbing potential*, ΔV, and the surface is warped outward. On either side of the region of mass excess, the plumbline is deflected inward. A mass deficiency has the opposite effect. For a single mass anomaly in an otherwise uniform earth:

$$N = \frac{\Delta V}{g} \tag{2.8-1}$$

where N, is the *geoidal undulation*, defined as the separation or height difference between the level ellipsoid and the geoid (Figure 2.8-1) and g is the mean value of gravity along N. The problem in the case of the actual earth is that there are a great number of mass anomalies contributing to the value of N at any point. In general, neither the masses, nor the disturbing potential, ΔV, are known, but the effect of the masses on g is.

2.9   STOKES' THEOREM

The difference, $\Delta g = g - \gamma_o$, between normal and measured gravity at a point on the geoid is called the *gravity anomaly*. The relation between N and the gravity anomalies, which is known as Stokes' theorem, was originally derived by Stokes in 1849. Stokes' theorem says, in essence, that at a point P on the geoid S, the geoidal undulation N, from a concentric reference figure with the same mass and volume as the geoid can be computed from Δg provided that (1) the gravity anomalies refer to the geoid, (2) S is an equipotential surface including all the topographic masses, and (3) the value of g has been measured everywhere over S. Stokes' formula is:

$$N = \frac{R}{4\pi\gamma} \int_S \Delta g \ S(\psi) \ d\sigma \qquad\qquad (2.9\text{-}1)$$

where R is the radius of a spherical approximation to the geoid, $S(\psi)$, is *Stokes' function*, ψ, is the angle between the radius at P and the radius to the variable element of surface over which the integration takes place, and $d\sigma$ is an element of solid angle. Stokes' function is given by:

$$S(\psi) = \mathrm{cosec}\ \frac{\psi}{2} - 6 \sin \frac{\psi}{2} + 1 - 5 \cos \psi - 3 \cos \psi \ \ln (\sin \frac{\psi}{2} + \sin^2 \frac{\psi}{2}) \quad (2.9\text{-}2)$$

Effectively, $S(\psi)$, is a weighting factor which weights the gravity anomalies according to their angular distance, ψ, from P. Its zeros are near $\psi = 39°$ and $\psi = 118°$. At the attracted point, $\psi = 0$, $S(\psi)$ becomes infinite, so that when evaluating N careful attention must be given to the neighborhood of the attracted point.

2.10   DEFLECTIONS OF THE VERTICAL

The angle between the normals to the level ellipsoid and the geoid at a point (Figure 2.8-1) is called the *deflection of the vertical*. The components of the deflection of the vertical in the north-south and east-west vertical planes, are equal to the slopes of the geoid, and therefore the derivatives of N in these directions ie.:

$$\eta = \frac{N}{X} \quad \text{and} \quad \xi = \frac{N}{Y} \qquad\qquad (2.10\text{-}1)$$

where X is towards north and Y is towards east.

2.11 REDUCTIONS OF GRAVITY OBSERVATIONS

Gravity on the continents is rarely measured at sealevel, and both geodetic and geophysical interpretations of the measurements require that a correction be made for height of the station. The variation of g due to a change in distance from the earth's centre is obtained immediately by differentiation. If we assume the earth to be spherical, and of mass, $M_\oplus$, the value of g at a point distant, r, from the centre is:

$$g = \frac{GM_\oplus}{r^2}$$

Hence:

$$\frac{g}{r} = -2\frac{GM_\oplus}{r^3} = -\frac{2g}{r} \qquad (2.11\text{-}1)$$

At sealevel this gives a vertical gradient of gravity of -3086 $nms^{-2}m^{-1}$. For most purposes, this value may be used anywhere on earth. A more complete evaluation takes into account the ellipsoidal shape of the earth.

Example 2.11-1

By *how much* does gravity decrease from the bottom to the top of a mountain 1000 m high? Take g = 9.8 ms^{-2} and $R_\oplus$ = 6400 km.

Solution

Using (2.11-1) gives:

$$\frac{g}{r} = -\frac{2g}{R_\oplus} = -\frac{2\times9.8}{6400\times1000} = -3060 \ nms^{-2}m^{-1}$$

Hence:

$$\Delta g = \frac{g}{r}\Delta h = -3060\times1000 = 3060 \ \mu ms^{-2}$$

SECOND   THIRD

FOURTH   FIFTH

EIGHTH   ELEVENTH
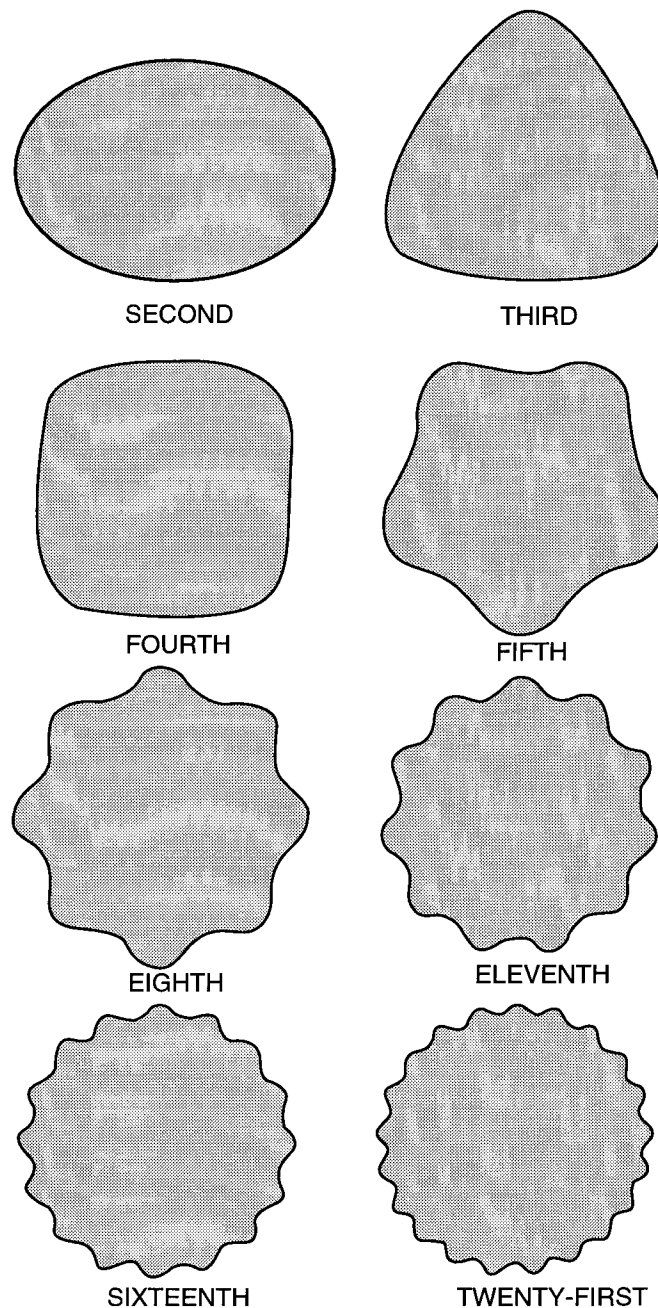
SIXTEENTH   TWENTY-FIRST

Figure 2.12-1
Harmonics and the earth's shape.

Equation (2.11-1) gives the rate at which g decreases with increasing distance from the earth's centre, or height, if no additional mass is interposed between the observer and the earth. The decrease is that which would be measured by an observer rising in a balloon through the air, and the correction is therefore called the *free-air correction*. If g is measured on the land surface at different heights, at the same latitude, the variation with height will not be that predicted by the free-air correction, because of the additional mass beneath the higher station. Bouguer, in 1749, realized this during the course of his measurements

in the Andes, and suggested that the additional attraction due to the material above sealevel be aproximated by treating this material as an infinite horizontal slab, of thickness equal to the height of the station. For a typical density of crustal material, the correction is 1118 $nms^{-2}m^{-1}$. The effect, which is known as the *Bouguer correction*, is of opposite sign to the free-air term.

2.12 HARMONICS OF GRAVITY AND THE SHAPE OF THE EARTH

The complex shape of the geoid can be built up by summing up the effects of what are called harmonics of the earth's gravity field. Let us assume for the moment that the earth is symmetrical around its axis and apply these harmonics to the basic sphere (Figure 2.12-1). More specifically, these particular harmonics are called *zonal harmonics* because they represent departures of the earth's shape from a sphere in zones parallel to the equator. The first harmonic is always zero if, as is customary, the earth's centre-of-mass is taken as being in the plane of the equator. The second harmonic represents a tendency toward an ellipse rather than a circle. This harmonic is another way of describing the basic ellipsoidal shape of the earth. Accordingly, the second harmonic is by far the largest of the series of harmonics. The third harmonic represents a tendency toward a triangle, often called a pear shape because if it were exaggerated the earth would look like a pear. The fourth harmonic is a square with smoothed corners. The fifth harmonic looks like a flower with five petals. The sixth harmonic with six petals, the seventh with seven, and so on, can easily be visualized. The even harmonics are symmetrical about the equator. Not so the odd harmonics. These give rise to different shapes in the northern and southern hemispheres. This picture may seem complicated enough, but it is in many ways greatly oversimplified. So far we have taken no account of the variations of the earth's shape with longitude, having assumed that the equator is an exact circle. Such variations, which are much smaller than the earth's flattening, can also be accounted for by harmonics.

REFERENCES

COOK, A.H., 1969. *Gravity and the Earth*, Wykeham, London, 95pp.

GARLAND, G.D., 1965. *The Earth's Shape and Gravity*, Pergamon, Oxford, 183pp.

HEISKANEN, W.A. & MORITZ, H., 1967. *Physical Geodesy*, W.H. Freeman, San Francisco, 364pp.

KING-HELE, D., 1967. The shape of the earth, *Scientific American*, **217**, 67-76.

TORGE, W., 1989. *Gravimetry*, Walter de Gruyter, Berlin, 465pp.

TORGE, W., 1991. *Geodesy* (Second Edition), Walter de Gruyter, Berlin, 264pp.

EXERCISES

2.1   Virtually all surveying measurements are influenced by the earth's gravity field. Discuss this statement using examples.

2.2   Gravity is not constant over the earth. Why?

2.3   Give the unit of gravity in the SI-system. Why is a smaller unit used in geodesy?

2.4   Show that the components of the force of gravity may be obtained from the potential of gravitation by differentiation.

2.5   What are equipotential surfaces and why are they important in studying the earth's shape?

2.6   Define gravity.

2.7   Define the gravity vector.

2.8   What are some of the temporal variations of gravity? Give their approximate magnitude.

2.9   Show that the plumbline is perpendicular to a level surface.

2.10   Give the relationship between potential difference, gravity and height.

2.11   The geoid is both a concept and a reality. Discuss this statement.

2.12   Define normal gravity.

2.13 What is a level ellipsoid?

2.14 Define the mean earth ellipsoid.

2.15 Why do we need gravity models?

2.16 What essentially are harmonics of gravity? How are they used to build up the shape of the geoid?

2.17 Define geoidal undulation. How is this used to define the shape of the geoid?

2.18 What is a gravity anomaly?

2.19 What assumptions are made when using Stokes' theorem?

2.20 Distinguish between free-air and Bouguer reductions of gravity measurements.

2.21 What are deflections of the vertical? Show that they depend on the shape of the reference ellipsoid.

3 Time

3.1 INTRODUCTION

The measurement of time is of fundamental importance in geodesy. This is demonstrated by the fact that time is the independent variable in the equations describing the position of a satellite as it orbits the earth. Time also forms the basis of all modern distance measuring devices used in surveying. Time is kept by clocks and, unlike other measurable quantities such as length, mass and temperature, cannot be held constant. A clock can be stopped; it then shows a point or instant on its own time scale, ie. the time when we stopped it. However, time goes on and if the clock we stopped happens to be the only one at hand, we would have lost a time scale forever. If we start the clock anew, it will be running late. By how much? This can only be found out with the help of another clock which has kept going during our experiment. Here we have implicitly used both meanings of the word time, namely time as *date* on a time coordinate axis having a defined origin, and time as *interval* during which the clock was stopped. In geodesy we use the term *epoch* to denote date (eg. the time at which a laser is fired to begin the satellite range measuring process) and *delay* to denote interval (eg. the time it takes for the transmitted photons to travel to the satellite and back).

3.2   FREQUENCY AND TIME

Frequency is a quantity closely related to time, sometimes very freely called its inverse. This is not strictly true but the meaning of frequency is derived from the observation of periodic events, ie. events which repeat at regular intervals of time, as for instance the oscillation of a pendulum. If constant, this interval is called the *period* ,T, and the frequency of oscillation, v, is indeed the inverse of the the period:

$$v = \frac{1}{T} \text{ (Hz)} \qquad (3.2\text{-}1)$$

Frequency is measured in *hertz* (Hz), one hertz being one period or one cycle per second. A particularly outstanding feature of frequency and time is the high accuracy of the basic definition and the precision of the measurements, which during the last 40 years or so have progressed to such a level that they leave all other measurements of physical quantities far behind.

3.3   THE SECOND

In the International System (SI) of units of measurement based on the convention of the metre, the fundamental unit is the second, the hertz being a derived unit. Traditionally, the definition of the second was based on astronomical concepts and observation and until 1956, the second was defined as 1/86 400 of the mean *solar day*. The solar day is the time between successive transits of the sun across the observer's meridian. Averaging this time out over a year gives the mean solar day.

As the irregularities of the earth's rotation had become well known since the early 1930s through the use of quartz crystal clocks as well as improved methods and instruments for astronomical observations, a definition allowing variations with time of the basic unit of measurement appeared no longer tolerable. Thus, in 1956 a new definition was adopted, based on ephemeris time (see Section 3.5), one second being 1/31 556 925.9747 of the tropical year. The *tropical year* is the time between successive arrivals of the sun at the first point of Aries (see Section 5.3). One tropical year is about 365.242 19 mean solar days. However, the ephemeris second was difficult to determine as several years of observation were required to reduce the measurement error to the required accuracy; a few parts in 10^9.

Meantime, new approaches to the problem of defining an invariant time unit were undertaken in the field of atomic and molecular spectroscopy. Taking advantage of the developments of microwave electronics before and during the second world war, physicists had discovered many atomic and molecular

spectral lines in the centimetre wave band. The use of a particular resonance as a possibly invariant frequency led to a new device, the *atomic clock*. The first of these devices was based on an absorption line at 23.87 GHz of the ammonia molecule. Its accuracy was limited to about 1 part in 10^7 and could therefore not yet compete successfully with traditional methods. However, subsequent experiments with other methods, especially the caesium atomic beam resonator led to rapid progress in the improvement of the accuracy of laboratory frequency standards. This rapid evolution produced a new definition of the second, which was no longer based on celestial mechanics but on quantum mechanics and in 1967, the second was redefined as follows:

"The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium atom 133".

Interestingly, frequency and not time is the measured quantity in this definition. Whilst in the old definition the second was given as a small fraction of a long period, we now have given the second as a large number of very rapid oscillations. This new definition had practical consequences on the operation of clocks.

3.4   TIME KEEPING

In the old practice, clocks were used to keep time ie. time was in the stars and the clocks helped to keep it between observations. Hours, minutes and seconds were the results obtained by dividing longer time periods such as the mean solar day or tropical year. With the new definition, longer time intervals are built up by successively adding elementary time intervals. Without changing their inner mechanisms clocks became time scale generators.

Clearly, *continuity* must be preserved between the traditional and new ways of defining and measuring time. There is a problem with our physical oscillator generating stable periodic oscillations - strict periodicity implies that each successive cycle is an exact copy of each preceding cycle. Hence, we can start counting at any time ie. there is no defined origin in our time coordinate. This problem is overcome by means of synchronization according to a conventional origin. This has been agreed upon, namely 1 January 1958, 0h, 0m, 0s.

Frequency Standards

In experimental work in science and engineering, a standard built and used for the purpose of physically realizing the defined unit is called a *primary standard*. Closest agreement with the defined unit is required only from a few primary laboratory standards. All other users can adjust the frequency of their clock oscillators to the primary standards. Accordingly, the important quality required

from these clocks is the *stability* of their frequency. The higher the stability of the oscillator, the better the uniformity of the general time scale.

Relativistic Effects

With the accuracies and stabilities currently achievable approaching a few parts in 10^{15}, the consideration of relativistic effects is necessary, especially for applications in space where high relative velocities and gravitational potential values different from those on the surface of the earth are encountered. Atomic clocks are currently the best available approximations to ideal clocks in the relativistic sense which define *Proper Time*. Now, in a set of clocks, spatially distributed within a given frame of reference, the time scale defined by these clocks referred to each other by a set of rules, is called *Coordinate Time.* Unfortunately, there is a danger of confusion with the term *coordinated time system* (see below) which means just a system of synchronized (in time) and adjusted (in frequency) clocks, not necessarily referred to a reference frame in the relativistic sense.

The international atomic time scale, TAI, the abbreviation for this and most other time scales reflecting the french order of the words (see later) is not only a coordinated time system but also a physical realization of coordinate time based on the SI-second as determined at sealevel. The TAI time scale is computed by Bureau International de Poids et Mesures (BIPM) in Paris as a weighted mean time scale based on the operation of many caesium clocks located at various laboratories around the globe. These clocks are now intercompared on a regular basis by means of GPS.

It is important to note that for using high precision time ordered systems, the theoretical basis is general relativity. This is especially true for distance measurements made in positioning or navigation since we assume the constancy and invariance of the speed of light, c, in vacuo and since distance measurements are in fact electromagnetic wave propagation measurements. We thus do not use metre sticks or even tapes but signal sources, clocks counters and the postulate that c is a universal constant.

Frequency Generators and Clocks

Oscillators provide frequency and, therefore, time in clocks. Examples of oscillators used in precision clocks are, (i) quartz crystals, (ii) rubidium vapour cells, (iii) caesium beam tubes and, (iv) hydrogen masers. The last three are atomic frequency standards based on the principles of quantum mechanics. We briefly describe these frequency generators below and review some of their general properties.

Quartz Crystal Oscillators. Quartz crystal oscillators, which are relatively inexpensive, compact and robust, are common devices used as stable frequency generators. The piezoelectric effect of quartz makes it easy to excite
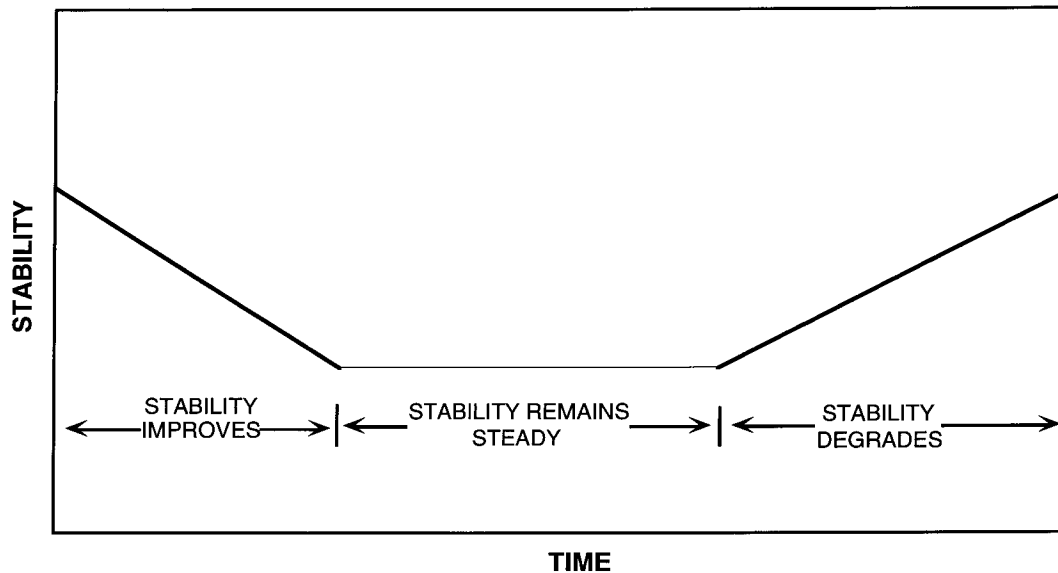
Figure 3.4-1
Stability *curve of an atomic oscillator (Allan variance).*

mechanical vibrations by means of an oscillating electric field on electrodes deposited on the surface of the crystal. A large variety of models exists, which are made for many applications, ranging from the subminiature low power units used in wrist watches to high quality oscillators enclosed in temperature-controlled ovens.

Rubidium Vapour Cells. The rubidium vapour cell oscillator is based on the 6 834 682 605 Hz resonance frequency of the hyperfine energy transition of Rb[87] atoms. At a cost of a few thousand dollars, they are the least expensive of atomic clocks

Caesium Beam Tubes. As already mentioned above, caesium beam tubes are based on the 9 192 631 770 Hz resonance frequency of the hyperfine energy transition of Cs[133] atoms and this is used to define the SI-second. They are the most accurate and most frequently used of the atomic clocks.

Hydrogen Masers. Hydrogen masers use the resonance frequency of a hyperfine energy of atomic hydrogen. They are the most stable of atomic clocks but also currently the most expensive. As such, they are used only in high precision applications of geodesy eg. very long baseline interferometry (see section 7.3).

Performance of Precision Oscillators

One measure which is used to describe the stability of a particular frequency standard is the *Allan variance.* Its precise definition leads to a lengthy discussion of statistics and is beyond the scope of this monograph. Suffice it
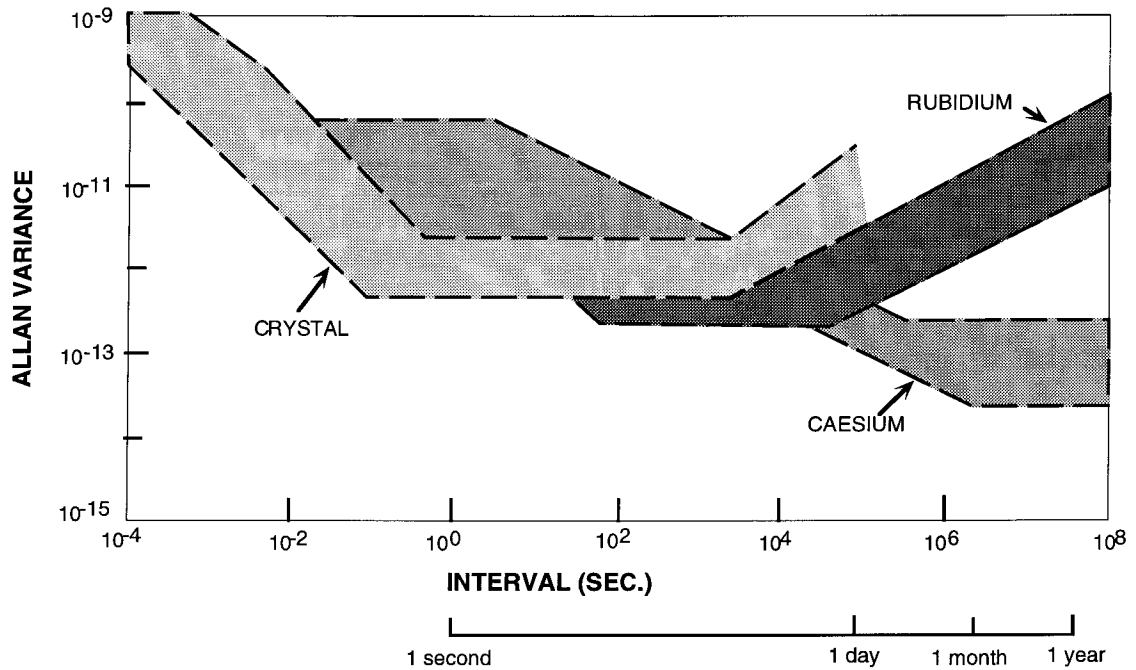
Figure 3.4-2
Allan variances for atomic oscillators.

here to say that it is a measure of precision which allows for the fact that most realworld oscillators exhibit both random frequency variations about an average as well as a systematic frequency drift with time. Nonetheless, the smaller the Allan variance the more stable the oscillator. The typical stability behaviour of a precision oscillator is shown in Figure 3.4-1. We observe that the stabilty at first increases as time elapses, then is steady for a while and subsequently drifts off. Stability is commonly expressed in parts of some power of 10 eg. 1 part in 10^{11}.

In comparing the performance of the various types of clocks or frequency standards a distinction must be made between the devices used in national standard laboratories and commercial instruments. National standards are operated by experts and are evaluated at regular intervals. However, they are not directly accessible by the great majority of users except through standard frequency and time disemination services. In a commercial or field environment, accuracy is not the concern; rather other factors, such as size, weight, environmental behaviour, power consumption and cost have an influence on the design. The ranges and time-domain stability measures of commercially available frequency standards is shown in Figure 3.4-2. The more conservative (upper bound) limit corresponds to manufacturers specifications. The best (lower bound) values correspond to measurements made on units in good laboratory conditions.

Example 3.4-1

The stabilty of a particular oscillator is 1 part in 10^{11}. *How many* years would it take for this oscillator to lose 1 second.

Solution

The oscillator loses 1 second every 10^{11} seconds. This is equivalent to:

$$\Delta T = \frac{1 \times 10^{11}}{24 \times 60 \times 60} \text{ days} = \frac{1 \times 10^{11}}{86400 \times 365} \text{ years} \approx 3200 \text{ years}$$

At present, caesium beam devices have an almost unique position in time keeping insofar as the principle is used not only in national standard laboratories as the primary reference but also in commercially available instruments. The difference is not in the principle but in the accuracy and intended use. Thus most caesium standards, whilst having the properties of primary standards within their proper accuracy limits, are operated as secondary standards of time and frequency. In these applications use is made of their long-term frequency stability which is superior to that of all other currently available oscillators.

Oscillator Reliability

Standard frequency generators used in clocks and timing systems must not only have the best possible long-term frequency stability but also be very *reliable* since any momentary failure immediately leads to the loss of the time scale. In times of rapidly evolving technology it is difficult to find the best compromise between high performance and reliabilty. Cost becomes the limiting factor if one attempts to improve reliabilty by redundancy ie. operating several devices simultaneously. Both caesium beam and rubidium vapour cell frequency standards seem to have about the same lifespan, while crystal oscillators last much longer.

3.5   TIME SCALES

Several time scales occur naturally in astronomy and geodesy. Before discussing these, we review some basic concepts of time scales. We leave aside philosophical concepts of time and considerations of gravitation and relativity theories, concerning ourselves mainly with physical time scales generated by clocks, especially in the form of a running display or sequence of signals which are used for dating of events.

Clocks and Time Scales

There is no readily available absolute time scale to be found in nature and an ideal device generating a perfect time scale is a purely theoretical construction. Thus, all real time scales based on either astronomical observations or on physical instruments are imperfect approximations to that theoretical concept. In a general sense, the concept of a clock, including the rotating earth and the movement of celestial bodies, involves three main parts and functions, namely: (1) a periodic movement which can be observed, (2) the continuous counting of the periods, and (3) the display of the registered count. Methods of performing these functions abound but whatever the method adopted, the result will be a particular time scale defined by that particular clock system ie. each clock defines its own time system. Among the periodic physical movements used to define practical time scales are, (i) free spinning rotors, (ii) Keplerian movement, and (iii) harmonic oscillators.

The earth is an example of a free spinning rotor and the rotation of the earth as observed by astronomical or geodetic means defines *Universal Time* (UT). The movement of a satellite around a central body based on some law of gravitation, such as the revolution of the earth around the sun and of the moon around the earth, are examples of time based on Keplerian movement (see Chapter 4). This is the basis of *Dynamical Time*. Most kinds of mechanical or electrical oscillations, including the oscillations of an electromagnetic wave (photon) emitted or absorbed by a quantum-mechanical system, belong to the class of harmonic oscillators. Examples are: pendulum, balance wheel with hairspring, tuning fork, quartz crystal and atomic resonators. The free spinning rotors and the Keplerian movement of celestial bodies form the basis of *Astronomical Time* and *Astronomical Time Scales.*, while the harmonic oscillators lead to the great variety of devices known as clocks in the popular sense. Among these clocks, we distinguish the atomic clocks as being the most accurate and this leads us the concept of *Atomic Time* and *Atomic Time Scales.*

Starting from any of the abovementioned periodic movements to establish a time scale by observation, continuous counting, registering and displaying, there are two main actions which have to be undertaken to obtain a time scale:

(1) the *period* or its inverse, the frequency of the basic oscillation must be measured, adjusted or defined; and

(2) the *origin* of the time scale must be specified.

The first action, which consists of establishing a unit of time interval, is not sufficient to define the time scale. Only the choice of origin from which we start counting the periods, completes the task. In practice, both actions require conventions to be agreed upon. Indeed, the origins of all practically used time scales are based on conventions obtained by international agreement. Whatever type of phenomenon is used to establish a time scale, the most important requirement is *uniformity*. Uniformity means that the intervals between

time scale marks are equal ie. constant period and frequency of the basic oscillation. Time scale uniformity and frequency stability are thus closely related.

Discussion

1. There are some differences in the practical implementation between astronomical time scales and atomic time scales. The duration of the period of the phenomena used in astronomical time keeping is very much longer than that of the oscillations used in laboratory clocks. In traditional astronomical time keeping, clocks are, therefore, needed for subdivision of the observation periods into smaller and more practical intervals such as hours, minutes and seconds. Also, the various astronomical phenomena have periods which are different and are not integer multiples of each other eg. earth rotation (1 day), lunar orbit (28 days), earth orbit (1 year). Furthermore, the very high frequency of oscillation occurring in atomic clocks allow the generation of time scales by counting only.

2. There has been some misunderstanding between astronomers and physicists in the years of atomic time keeping, most of it due to a conflict between tradition and some disrupting new ideas. In the old days, the concept of a time scale was not used. There was time and its passage had to be determined by observing the stars. Clocks were neither accurate nor reliable and the idea of relying on clocks to define a time scale was thought to be risky and unwise. Atomic clocks were accepted for defining better time intervals but not time itself. What would happen if all clocks stopped or were destroyed? In such a case the time scale defined by these clocks would be lost forever. In practice the likelihood of this occurring has been overcome by redundancy and wide geographical distribution of time-keeping institutes. Astronomical observations have not been discontinued either. Therefore, the risk of losing time due to such a cataclysm is rather remote.

3. Today, there is no longer any controversy as both types of time scales are continuously being kept and scientists working in positional astronomy, celestial mechanics, geodesy etc. no longer dispute the convenience of accurate and reliable time keeping based on atomic clocks. Also, the angular position of the rotating earth with reference to the sun and the stars continues to be required for geodesy, navigation and civil time keeping (day, night, noon etc.).

Clearly, several time scales are required. We give a short description of the more important ones below.

Earth Rotation Time

The following time scales are based on the rotation of the earth:

Sidereal Time. A sidereal day is the interval between two transits of the same star through the local meridian ie. the duration of one period of rotation with respect to the system of fixed stars.

Mean Solar Time. Actual solar time as determined from the position of the sun in the sky is not uniform throughout the year because of the nonuniform motion of the earth around the sun (Kepler's second law). By averaging over one year, Mean Solar Time is obtained. This has a known relation to sidereal time, the ratio of the sidereal day to the mean solar day being roughly 1.00274 ie. one sidereal day is approximately 23 h 56 m.

Universal Time. Sidereal time at a given location on the earth converted to mean solar time and referred to the Greenwich meridian is called UT0. Applying corrections for the position of the pole (see Section 6.6) results in UT1 which is the same at all points of the earth. UT1 gives the angular position of the earth relative to the stars. Applying corrections for seasonal variations of the speed of rotation of the earth results in UT2. However, there are unpredictable variations in the rotation of the earth and over long time intervals UT2 is not significantly more uniform than UT1. The term *Greenwich Mean Time* (GMT) is still popular. This should be replaced by UT1.

Dynamical and Ephemeris Time

Ephemeris Time (ET) was the uniform time scale used to determine the position of celestial bodies. The scale was defined by the orbital motion of the earth about the sun. As already mentioned above, the second of ET was defined as 1/31 556 923 9747 of the tropical year for 0 January 1900. However, in practice, ET was obtained from the orbital motion of the moon around the earth. Since 1976, two new dynamical time scales, *Barycentric Dynamical Time* (TDB) and *Terrestrial Dynamical Time* (TDT), which are based on the SI-second have been used. A clock fixed on the earth exhibits periodic variations as large as 1.6 msec with respect to TDB due to the motion of the earth in the sun's gravitational field. However, in describing the orbital motion of near-earth satellites we need not use TDB, nor account for these relativistic variations, since the effect is the same on both the earth and the satellite. For satellite computations we use TDT which represents a uniform time scale for motion within the earth's gravity field and which has the same rate as that of an atomic clock on earth. In the language of general relativity TDB corresponds to *Coordinate Time* and TDT to *Proper Time*. TDT is aligned with the epoch of TAI (see below) on 1 January 1977. ET, as such, is now only of historical interest. TDT is defined by an offset from TAI, specifically:
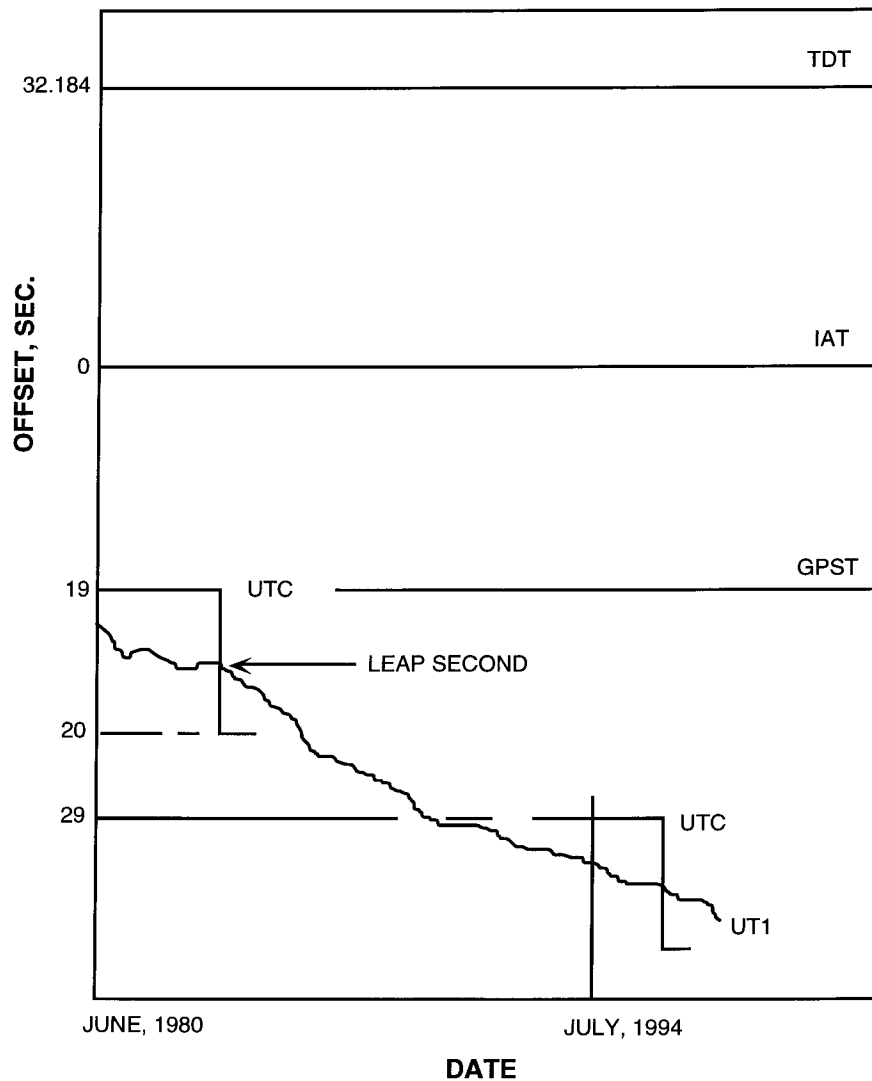
TDT = TAI + 32.184 s

Figure 3.5-1
Generalized relationship between time scales.

Atomic Time

Atomic Time is a time scale obtained by continuous counting of SI seconds. By international agreement, the origin of atomic time scales has been set at 1 January 1958, at 0h 0m 0s UT2. Since each clock defines its own time scale and no clock is perfect, the time scales initially synchronized on this conventional origin will depart from each other as time goes by. International coordination of timekeeping activities is needed to reach a common recognized time scale which is uniform. The institution charged with the task of establishing the *International Atomic Time Scale* (TAI) is the BIPM. The fundamental unit of TAI (and therefore TDT) is the SI-second defined above. The SI-day is defined as 86 400 SI-seconds.

Coordinated Universal Time (UTC). Despite its name, UTC is a time scale generated by atomic clocks. The continuing requirement for a time scale approximating UT is due to its wide application in surveying and navigation. A compromise solution had, therefore, to be found which retains the advantage of uniform time scale generation by atomic clocks and still follows the variations of the earth's rotation. UTC is kept in synchronization with UT1 in such a way that second jumps, or *leap seconds*, are introduced in UTC at the end of the last day of a month when UT1-UTC exceeds 0.7 seconds. The earth rotation rate is currently decreasing by about 1 second/year and this adjustment is made accordingly. As a consequence of the irregularity of the earth's rotation, the relationship between UT1 and UTC must be defined by observation. Also, UTC differs from TAI by an integer number of seconds. UTC and TAI were identical in 1962. Currently (June, 1996):

$$UTC - TAI = 30 \text{ s}$$

Each of the world's sites contributing to the BIPM keeps a local realization of UTC, the offset and rate of which relative to UTC(BIPM) are monitored and periodically corrected. In Australia, UTC(Australia) is maintained by AUSLIG in Canberra and the relationship of UTC(Australia) to UTC(BIPM) is known to within a few μsec.

GPS Time (GPST). This is time determined by atomic clocks at the GPS master control site near Colorado Springs, Colorado, USA. These clocks are synchronized periodically with UTC and GPST was set to UTC at 0 hours on 6 January 1980. Thus, there is a constant offset of 19 seconds between GPST and TAI, that is, at any instant:

$$GPST = TAI - 19 \text{ s}$$

The GPS satellites broadcast GPST every second.

Figure 3.5-1 depicts the generalized relationships between the various time scales. All clocks are assumed to run at the same rate as IAT clocks.

Example 3.5-1

A satellite is observed at epoch MJD = 48 774^d.333 333 333 333 UTC. *Given* that UT1-UTC and UT1-IAT at 0^h UTC for MJD 48 774 and MJD 48 779 respectively are -0.50810, -28.50810, -0.51690 and -28.51690 seconds, *determine* UT1, TAI, TDT and GPST in years, days, hours, minutes and seconds.

Solution

The *Julian Day Number* is the number of the day in a consecutive count beginning as far back in time so that every date in the historical era can be

TABLE 3.5-1 *Julian Day Number of Day Commencing at Greenwich Noon on:*

1992-1999

Year	Jan 0		Feb 0	Mar 0	Apr 0	May 0	Jun 0	Jul 0	Aug 0	Sep 0	Oct 0	Nov 0	Dec 0
1992	244	8622	8653	8682	8713	8743	8774	8804	8835	8866	8896	8927	8953
1993		8988	9019	9047	9078	9108	9139	9169	9200	9231	9261	9292	9322
1994		9353	9384	9412	9443	9473	9504	9534	9565	9596	9626	9657	9687
1995	244	9718	9749	9777	9808	9838	9869	9899	9930	9961	9991	*0022	*0052
1996	245	0083	0114	0143	0174	0204	0235	0265	0296	0327	0357	0388	0418
1997		0449	0480	0508	0539	0569	0600	0630	0661	0692	0722	0753	0783
1998		0814	0845	0873	0904	0934	0965	0995	1026	1057	1087	1118	1148
1999	245	1179	1210	1238	1269	1299	1330	1360	1391	1422	1452	1483	1513

included as a positive number. This uninterrupted series of days began at Greenwich mean noon on 1 January 4713 BC in the Julian Proleptic Calendar. The *Julian Date* (JD) at this instant was 0. The *Modified Julian Date* (MJD), where MJD = JD - 2 400 000.5, is used in modern work.

The Julian Day Numbers at Greenwich Mean noon for 1992 - 1999 are shown in Table 3.5-1. This table tells us that the satellite was sighted on 1 June 1992 at 08^h 00^m $0^s.00$ UTC. In order to determine UT1 and IAT at this time we need to interpolate the given values of UT1-UTC and UT1-IAT:

$$\Delta(UT1\text{-}UTC) = \Delta(UT1\text{-}TAI) = \frac{-0.00880}{5} \times 0.3333 = -0.000\ 59\ s$$

Hence:

UT1 = UTC - 0.50869 s = 07^h 59^m $59^s.49131$

IAT = UTC - 28 s = 07^h 59^m $32^s.0$

TDT = IAT + 32.184 s = 08^h 00^m $04^s.184$

GPST = IAT - 19 s = 07^h 59^m $13^s.0$

Example 3.5-2

What is the Julian Date corresponding to a UT of 1 January, 1995 at 06^h 48^m?

Solution

Since the Julian Date is measured from noon to noon, then midnight, or exactly 1 January, can be thought of as occurring 12 hours after 31 December, Julian Date. Hence from Table 3.5-1, for 31 December (0 January):

$$JD_{NOON} = 2\ 449\ 718.0$$

To this number we add 12 hours or 0.5 days, so that:

$$JD_{MIDNIGHT} = 2\ 449\ 718.5$$

and, finally $06^h\ 48^m$ is converted into days:

$$\Delta = \frac{6}{24} + \frac{48}{60 \times 24} = 0.283\ 333$$

which, when added to 1 January JD, yields the Julian Date at the required instant:

$$JD = 2\ 449\ 718.783\ 333$$

REFERENCES

DANA, P.H. & PENROD, B.M., 1990. The role of GPS in time and frequency dissemination, *GPS World*, July/August, 38-43.

GOUZHVA, Y.G., GEVORKYAN, A.G., BOGDANOV, P.P. *et al.*, 1995. Getting in sync: GLONASS clock synchronization, *GPS World*, April, 48-56.

HELLWIG, H., 1974. Frequency Standards and Clocks: a Tutorial Introduction, *NBS Tech. Note 616*, US Dept. of Commerce, Washington, DC, 64pp.

KARTASCHOFF, P., 1978. *Frequency and Time*, Academic Press, London, 260pp.

KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A. & COLLINS, J., 1987. *Surveying with the Global Positioning System*, F. Dümmler, Bonn, 128pp.

LANGLEY, R., 1991. Time, clocks and GPS, *GPS World*, November/December, 38-42.

WAHR, J., 1986. The earth's inconstant rotation, *Sky & Telescope*, June, 545-549.

EXERCISES

3.1   Time is used in what two basic contexts in modern positioning methods?

3.2   The definition of the second has changed over time. Explain why.

3.3   Define the SI second.

3.4   State the ideal requirements of an atomic clock and explain what these requirements mean.

3.5   Give a graph showing how the stability of an atomic oscillator varies with time. Hence distinguish between the stabilities of crystal, rubidium, caesium and hydrogen maser oscillators.

3.6   What basically is an Allan variance?

3.7   The stability of a particular oscillator is 1 part in 10^{14}. How many years would it take to lose 1 second?

3.8   What is a time scale? Why are caesium oscillators used to define modern time scales?

3.9   Name three atomic time scales used in geodesy.

3.10   Distinguish between the solar and sidereal day. Why is one longer than the other?

3.11   The earth is not a good time keepers. Discuss this statement.

3.12   Distinguish between dynamical, sidereal and atomic time scales. Why do we need all of them?

3.13   Using a diagram to illustrate your answer give the relationship between IAT, TDT, GPST, UTC and UT1 in a general way.

3.14   What are Julian Day numbers? Hence define Modified Julian Date.

3.15   Show that 1 millisecond of time represents a position change of approximately 45 cm at the equator.

4 Earth-Satellite Motion

4.1 INTRODUCTION

Satellites contribute to geodesy in various ways, (i) as a target for determining the positions of tracking stations and, by repeating the observations over time, to determine if relative movements of these stations occur, (ii) as a measure of the forces acting on the satellite, predominantly gravity, and (iii) as a reference frame with respect to which the motion of the earth as a whole can be measured. The potential of satellites for geodesy was recognized in the late 1950s when J. O'Keefe suggested that they could be used as elevated targets to strengthen geodetic networks on continental scales and to connect geodetic systems on ocean-separated continents into a single uniform geodetic system which could be used for precise mapping and navigation. The use of satellites for determining the earth's shape and gravity field was also recognized around this time and the first results for the earth's flattening were published soon after the launch of the Sputnik spacecraft. The early tracking of satellites for geodetic purposes was carried out using the Baker-Nunn camera network and these instruments continued to provide important data through the 1960s. Valuable geodetic data was also provided by Doppler tracking of the TRANSIT satellites. Both these Doppler and camera observations provided positional accuracies that were of the order of 10-20 m. Now the satellites are tracked with accuracies of a few centimetres and better, using lasers and radar methods, and positioning accuracies have improved by nearly three orders of magnitude.

Here we give the basic equations which describe the motion of a satellite about a parent body eg. an artificial satellite and the moon relative to the earth, or a

planet relative to the sun. We begin with Kepler's laws of planetary motion which, to a first order, also apply to artificial earth-satellites. We then define the parameters or elements, called Keplerian elements after Johann Kepler, which are commonly used to describe the geometry of an orbit and to relate the orbit to a space-fixed reference frame. Finally, we discuss the various disturbing effects which cause the orbit to depart from the ideal described by Kepler's laws.

4.2 KEPLER'S LAWS AND GRAVITATION

The basic laws which govern the motion of a satellite around a parent body were discovered by Kepler. Using data meticulously collected by Tycho Brahe, Kepler hit upon the ellipse as a possible solution of the various observed positions of the planet Mars. From this, Kepler published his first two laws of planetary motion in 1609. The third law followed about 10 years later. Kepler's laws are:

1. The orbit of each planet relative to the sun lies in a fixed plane containing the sun, and is an ellipse of which the sun occupies a focus.

2. The line joining the sun to each planet sweeps out equal areas of its ellipse in equal times.

3. The squares of the orbital periods of the planets are proportional to the cubes of their mean distances from the sun.

The first law describes the geometry of the orbit (ie. the orbit is an ellipse), while the second and third laws define the motion of the planet in the orbital plane. Kepler discovered them as unexplained facts. The full dynamical implications of these laws were only seen about 100 years later by Isaac Newton after he had formulated his laws of motion leading him to discover the law of universal gravitation.

4.3 TWO-BODY ORBITAL MOTION

Let us examine the motion of a body relative to a parent body. This is called the *two-body problem* in celestial mechanics. We use Newton's second law of motion and his law of universal gravitation. Moreover, we make the following simplifying assumptions:

- the bodies are spherically symmetric, enabling us to treat the bodies as point masses.
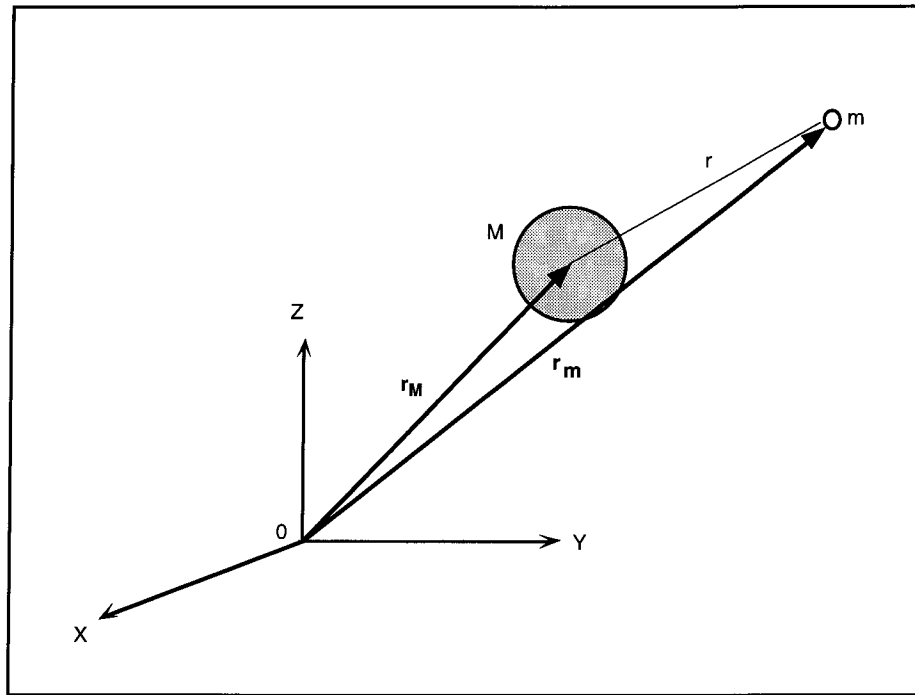
Figure 4.3-1
Two-body orbital motion.

- there are no external forces acting other than gravitational forces, which act along the line joining the centres of the bodies.

Newton's laws of motion are valid only in an unaccelerated and nonrotating (or *inertial*) reference frame. Newton described this inertial reference frame as being fixed in absolute space. However, he failed to indicate how one found this frame which was absolutely at rest. We return to this problem in Chapter 6. Here we shall assume that such an inertial frame exists and has been found.

Consider a system of two bodies of mass M and m (Figure 4.3-1). Let X,Y,Z be an inertial set of Cartesian coordinate axes with origin at the point O. The position vectors of the bodies M and m relative to the X,Y,Z-system are r_M and r_m respectively ie.:

$$r = r_m - r_M \qquad (4.3-1)$$

Applying Newton's laws in the inertial frame X,Y,Z and using a double dot above the r to denote double differentiation w.r.t time we have:

$$m\ddot{r}_m = -\frac{GMm}{r^2}\frac{r}{r} \qquad (4.3-2)$$

$$M\ddot{r}_M = \frac{GMm}{r^2}\frac{r}{r} \qquad (4.3-3)$$

or,

$$\ddot{r}_m = -\frac{GM}{r^3} \, r \tag{4.3-4}$$

$$\ddot{r}_M = \frac{Gm}{r^3} \, r \tag{4.3-5}$$

Subtracting (4.3-5) from (4.3-4) gives the vector differential equation of relative motion of the two bodies:

$$\ddot{r} = \ddot{r}_m - \ddot{r}_M = -\frac{G(M+m)}{r^3} \, r \tag{4.3-6}$$

We are particularly interested in the motion of an artificial earth-satellite, for which $m \ll M_\oplus$, where $M_\oplus$ is the mass of the earth. Hence, we may write:

$$G(M_\oplus + m) \approx GM_\oplus \tag{4.3-7}$$

and the equations of motion become:

$$\ddot{r} + \frac{GM_\oplus}{r^3} \, r = 0 \tag{4.3-8}$$

Equation (4.3-8) is a second-order vector differential equation in time ie. there are in fact three equations, one for each component of acceleration. Integration produces the satellite trajectory or orbit. This integration is not routine and we state without proof:

- the family of curves called conic sections (circle, ellipse, parabola and hyperbola) represent the only possible paths for an object in the two-body problem.

- the prime focus of the conic orbit is located at the centre of the central body (eg. earth and sun).

- the sum of the kinetic and potential energy of the satelllite stays constant, implying that the satellite must slow down or speed up as it moves around in its the orbit.

-the orbital motion takes place in a plane which is fixed in inertial space.
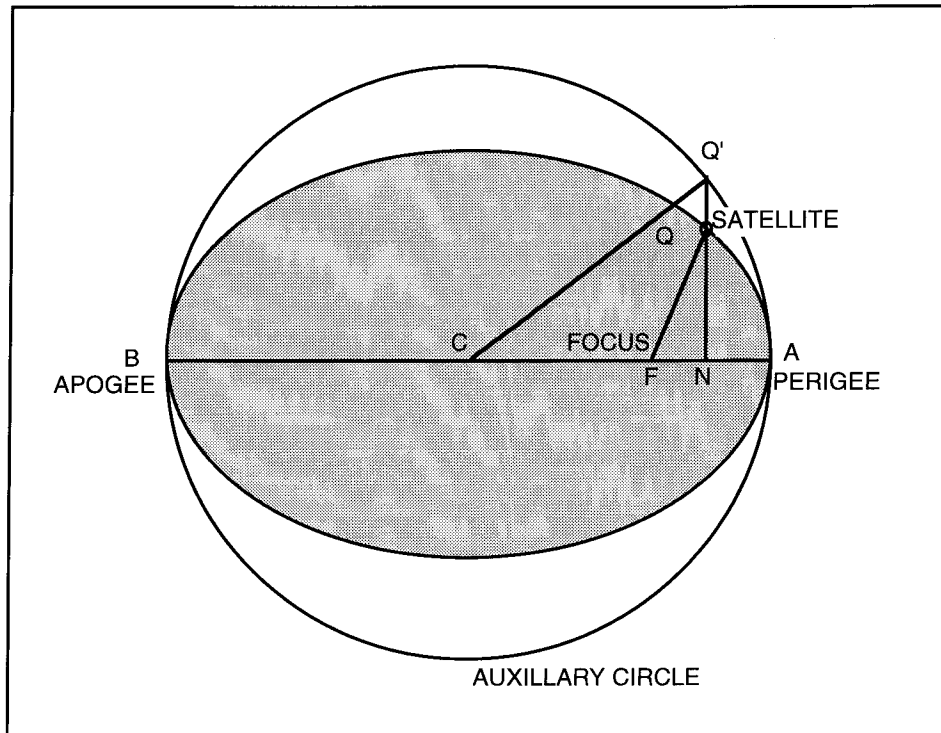
Figure 4.4-1
Geometry of orbital ellipse

4.4   ELLIPTIC MOTION

We are especially interested in the orbits of artificial earth-satellites which are ellipses. In Figure 4.4-1, the orbital ellipse AQB whose centre is at C touches at *perigee*, A, and at *apogee*, B, a concentric circle C whose radius is, a, the semi-major axis of the ellipse. The circle C is known as the *auxillary circle*, and is related to the ellipse by the equation:

$$QN = Q'N \sqrt{1-e^2}$$

Angle AFQ = v, is the *true anomaly*, where F is the occupied focus or *prime focus* of the ellipse; the angle ACQ' = E is known as the *eccentric anomaly*. The distance CF = ea.

Now:

$$r^2 = (FQ)^2 = (FN)^2 + (QN)^2 = a^2(\cos E - e)^2 + a^2(1-e^2)\sin^2 E$$

So:

$$r = a(1-e\cos E) = \frac{a(1-e^2)}{1+e\cos v} \qquad (4.4\text{-}1)$$

which is the equation of the orbit in polar coordinates r and v.

Further:

$$\sin v = QN/r = \sqrt{1-e^2}\ \sin E/(1-e\cos E)$$

$$\cos v = FN/r = (\cos E - e)/(1-e\cos E)$$

which gives:

$$\tan \frac{v}{2} = \sqrt{\frac{1+e}{1-e}}\ \tan \frac{E}{2} \qquad (4.4\text{-}2)$$

Equation (4.4-2) allows us to compute the true anomaly, v, from the eccentric anomaly, E, and vice versa.

The *mean motion*, n, in an elliptical orbit is defined as $2\pi/P$, where P is the *orbital period*. The mean motion may be computed from Kepler's third law:

$$n^2 a^3 = GM_\oplus \qquad (4.4\text{-}3)$$

Let T be the *time of perigee passage* and t the time; t-T is the time since perigee passage; then (t-T)/P is the fraction of the ellipse's area described by FQ since perigee and this fraction is the same as $n(t-T)/2\pi$. The quantity:

$$M = n(t-T) \qquad (4.4\text{-}4)$$

is called the *mean anomaly*. But this fraction, the area AFQ divided by the area of the ellipse, is the same as the area AFQ' divided by that of the area of the circle, and AFQ' is the area of the circular sector ACQ' less the area of the triangle CFQ'. Thus:

$$\frac{M}{2\pi} = \frac{Ea^2 - ea^2\sin E}{2\pi a^2}$$

Hence:

$$M = E - e\sin E \qquad (4.4\text{-}5)$$

which is known as *Kepler's equation*. Equations (4.4-4) and (4.4-5) allow us to compute the mean anomaly from the time of perigeee passage, T, or the eccentric anomaly, E.

Two other terms frequently used to describe orbital motion are *direct* and *retrograde*. Direct motion means easterly. This is the direction in which the sun, earth and most of the planets and their moons rotate on their axes and the direction in which all of the planets revolve around the sun. Retrograde motion is in the opposite direction. Orbital inclinations between 0° and 90° imply direct orbits and inclinations betwen 90° and 180° imply retrograde orbits.

Note

The complete solution of the problem of elliptic motion is contained in the equations which have been given. Thus, suppose that $GM_\oplus$, a, e and T are known. Then n can be found from (4.4-3). If the position of one object relative to the other is desired at some specified time t, (4.4-4) furnishes M and then (4.4-5) permits E to be found, from which by (4.4-1) and (4.4-2) one finds r and v.

Example 4.4-1

For a satellite in an elliptical orbit around the earth, a = 12.2700×10^6 m, e = 0.004 44 and T = 5 May 1976, $21^h \; 40^m 33$ UT. *Assuming* that $GM_\oplus = 3.9860 \times 10^{14}$ $m^3 s^{-2}$, *determine* M, E, v and r on 6 May 1976, $0^h \; 00^m 00$ UT.

Solution

We first calculate the *mean motion*, n, of the satellite using (4.4-3):

$$n = \sqrt{\frac{GM_\oplus}{a^3}} = \sqrt{\frac{3.9860 \times 10^{14}}{(12.27 \times 10^6)^3}} = 4.65 \times 10^{-4} \text{ rad. s}^{-1}$$

For interest sake, we determine the *orbital period*, P:

$$P = 2\pi/n = 225 \text{ min.}$$

which is to say, that the satellite orbits the earth once every 225 minutes.

With t-T = 8380 s ($2^h \; 19^m 67$), we obtain the *mean anomaly*, M, from (4.4-4):

$$M = n(t-T) = 4.65 \times 10^{-4} \times 8380 \times 180/\pi = 223°26$$

Kepler's equation (4.4-5) gives the *eccentric anomaly*, E:

$$E = M + e\sin E$$

In this formula the angles M and E should be expressed in radians. The calculations must thus be performed in the "radian mode" on our calculator. This

can be avoided by multiplying e in (4.4-5) by 180/π, the conversion factor from degrees to radians. Let e_0 be the thus modified eccentricity. Kepler's equation is then:

$$E = M + e_0 sinE$$

and we may now calculate in degrees.

To solve the above equation, we give an approximate value to E in the right side of the formula. Then the formula will give a better approximation for E. This is repeated until the required accuracy is obtained, here say $0°.0001$. For a first approximation, we use E = M. We thus have:

$$E_0 = M = 223°.26$$

$$E_1 = M + e_0 sinE_0 = 223°.0856$$

$$E_2 = M + e_0 sinE_1 = 223°.0862$$

$$E_3 = M + e_0 sinE_2 = 223°.0862$$

which has converged to the set tolerance level. Hence:

$$E = 223°.08$$

Equation (4.4-2) relates the *true* and *eccentric anomalies*:

$$tan\frac{v}{2} = \sqrt{\frac{1+e}{1-e}} tan\frac{E}{2} = \sqrt{\frac{1+0.00444}{1-0.00444}} tan\ 111°.54$$

Since v and E must lie in the same quadrant:

$$v = 222°.91$$

Finally, (4.4-1) gives the *orbital radius*, r:

$$r = \frac{a(1-e^2)}{1+ecosv} = a(1-ecosE) = 12.3098 \times 10^6\ m$$

Example 4.4-2

The period of an artificial satellite of the earth is 92.5 min and the semi-major axis of its orbit is 6800 km. *Given* that G = 6.67×10^{-11} $m^3 kg^{-1} s^{-2}$, *calculate* the mass of the earth?

If the mass of the moon is 1/81 times that of the earth, *what* is the period of a satellite about the moon in an orbit of semi-major axis 2000 km?

Solution

The *mean motion* is:

$$n = 2\pi/P = \frac{2\pi}{92.5} = 6.8 \times 10^{-2} \text{ rad. min}^{-1}$$

Manipulation of (4.4-3) gives:

$$M_\oplus = \frac{n^2 a^3}{G} = 6 \times 10^{24} \text{ kg}$$

Thus:

$$M_M = M_\oplus/81 = 7.4 \times 10^{22} \text{ kg}$$

Equation (4.3-3) gives:

$$n = \sqrt{\frac{G M_M}{a^3}} = 7.86 \times 10^{-4} \text{ rad. s}^{-1}$$

So, the *orbital period* is:

$$P = 2\pi/n = 133 \text{ min.}$$

4.5   THE ORBIT IN SPACE

It is usual to employ earth-centred (*geocentric*) equatorial axes OX, OY, OZ to describe bodies that orbit about the earth; a sun-centred (heliocentric) ecliptic system is used for bodies which describe orbits about the sun or for the lunar orbit. The *ecliptic* is the plane of the earth's orbit about the sun . In Figure 4.5-1, the origin O, is the geocentre; XOY is the plane of the equator; OX points towards a particular position of the earth in the ecliptic plane, known as the *vernal equinox* and denoted by the symbol, Υ (see Section 5.3); and OZ points towards the mean north pole (see Section 5.4). We return to this reference frame in Chapter 5. The satellite's orbit plane intersects the equatorial plane in the *line of nodes* ON, and if N is where the object crosses the equator from south to north, it is called the *ascending node*.
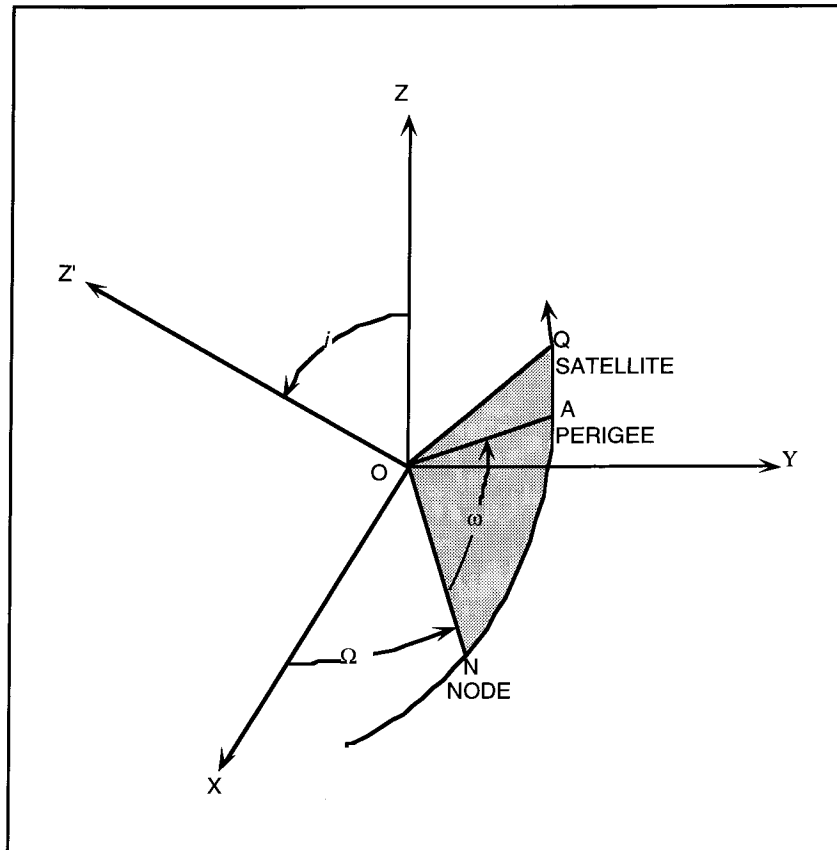
Figure 4.5-1
Orbit in space.

The angle XON is the *longitude* or *right ascension of the ascending node*, and is denoted by, Ω. Also, if OZ' is normal to the orbit, in which the object's motion appears clockwise to someone looking along OZ', then ZOZ' is the *inclination i* of the orbit. Thus the two angles Ω and *i*, specify the orientation of the orbit plane in space. The angle NOA, measured in the direction of the orbital motion, defines the position of perigee, A, and is denoted by, ω. It is called the *argument of perigee* and specifies the orientation of the orbit in its plane. The elements a and e for an ellipse, specify the orbit's shape and size. Thus the six elements which can be regarded as describing an orbit about a parent body are a, e, i, Ω, ω and M. The elements T, E and ν can be used instead of M but M is generally preferred in artificial satellite problems. These six elements correspond to the six initial data, three velocities and three coordinates of the satellite, that determine the orbit dynamically. This is often called the *state-vector* of the satellite.
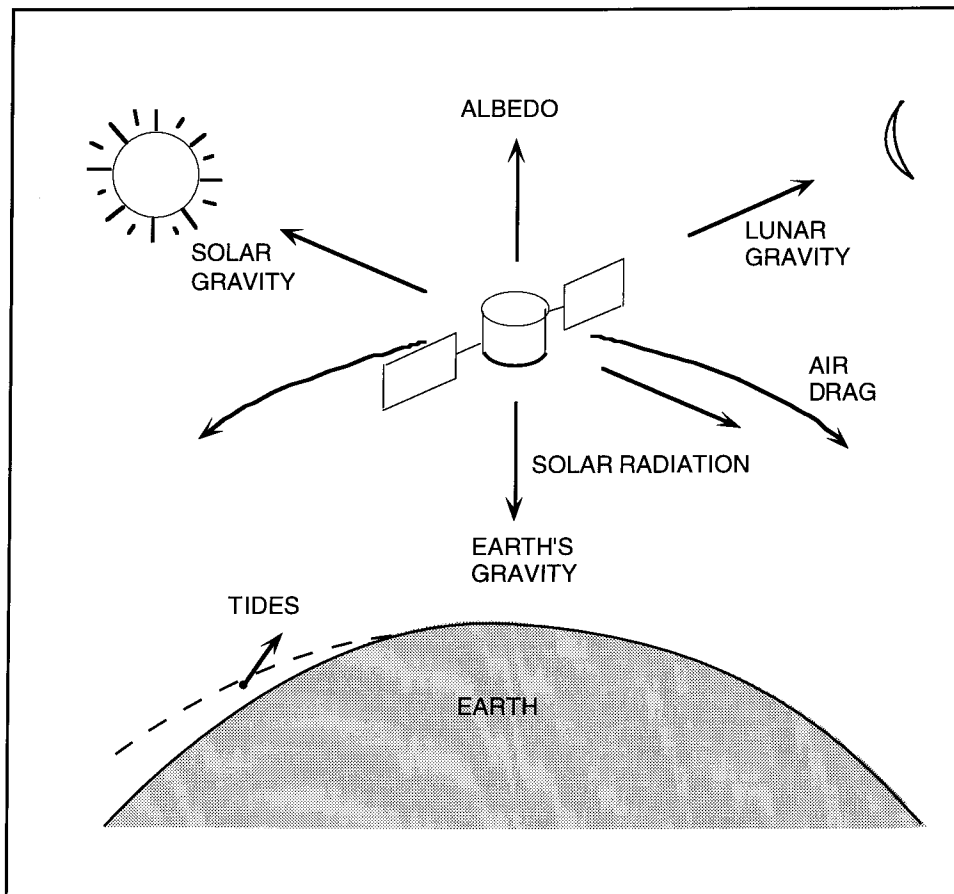
Figure 4.6-1
Main perturbing forces acting on a near-earth satellite.

4.6   PERTURBED ORBITAL MOTION

The actual orbit of a near-earth satellite departs from the Keplerian orbit due the effects of various perturbing accelerations of gravitational and nongravitational origin. These include the nonsphericity of the earth, the attraction of additional bodies such as the moon, sun and planets, atmospheric drag, direct and reflected (albedo) solar radiation pressure and, earth and ocean tides (Figure 4.6-1). Nonetheless, the orbit at any instant remains an ellipse which is described by the current or *osculating* orbital elements a, e, *i*, Ω, ω and M. However, the orbital elements, and thus the shape, size and space orientation of the ellipse, change with time. The equations of perturbed motion of the satellite about the earth become:

$$\ddot{r} = -\frac{GM_\oplus}{r^3} r + \Delta\ddot{r} \qquad\qquad (4.6\text{-}1)$$
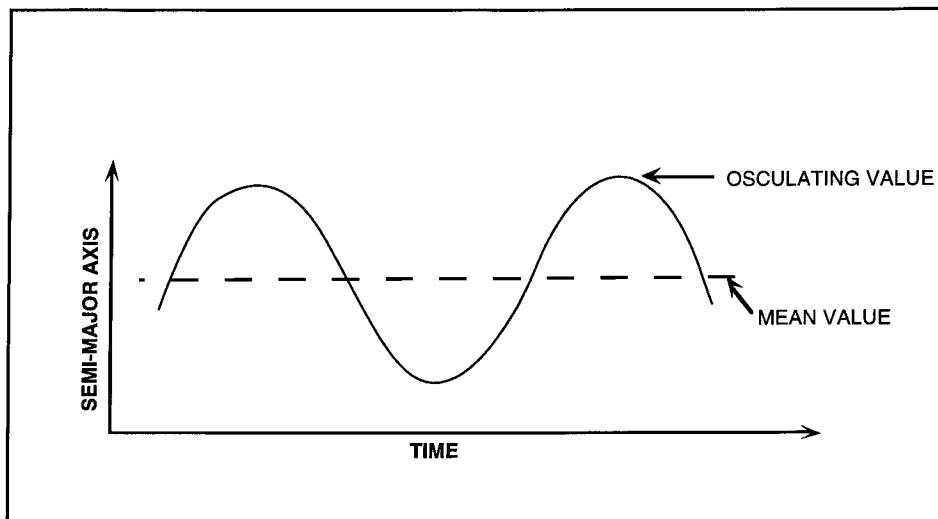
Figure 4.6-2
Perturbations of the semi-major axis due to gravity variations.
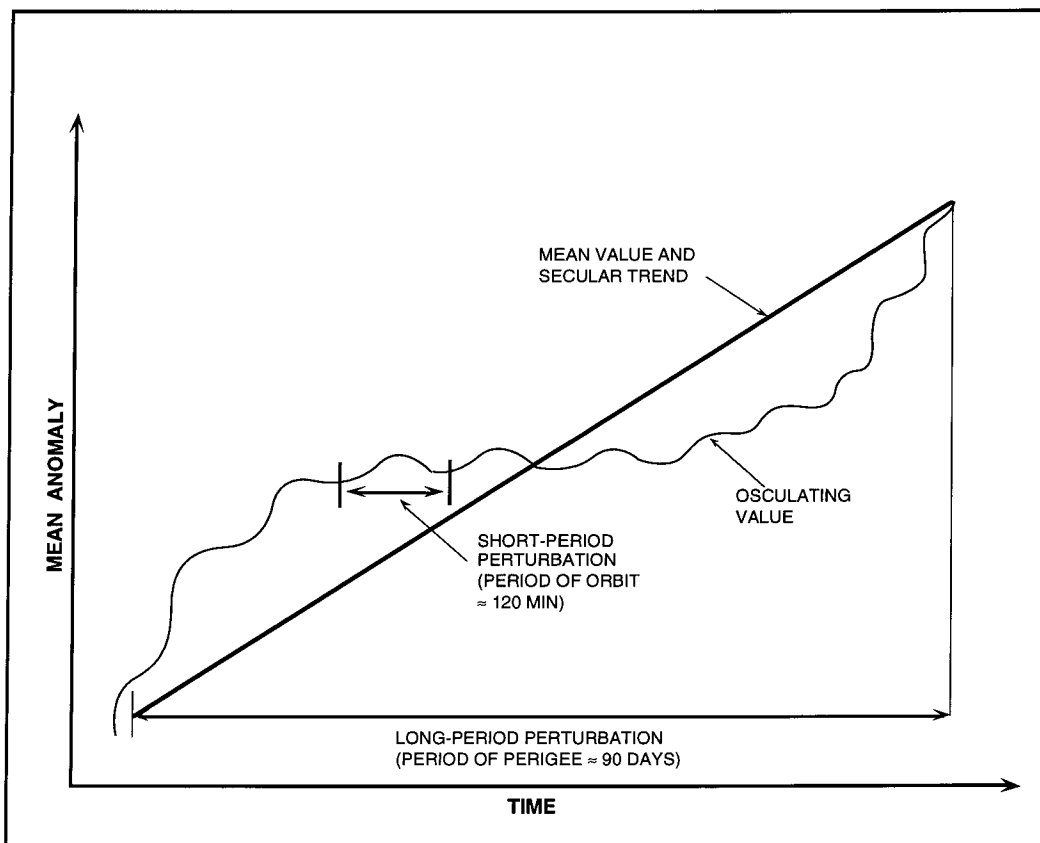
Figure 4.6-3
Short-period, long-period and secular perturbations of the mean anomaly due to gravity variations.

where the first term represents the two-body or central acceleration and the second term is the sum of the perturbing accelerations.

From observations it has been noted that the osculating Keplerian elements oscillate about some average value. For example, the perturbation of the semi-major axis caused by gravity perturbation of the orbit is shown in Figure 4.6-2. The perigee, node and mean anomaly have a more complex motion. They involve *secular* (varying linearly with time), *long-period* and *short-period* perturbations as shown in Figure 4.6-3. Short-period perturbations usually have a period of approximately the period of the satellite, while long-period terms have a period proportional to the period of perigee or even longer. Below we summarize the effects of the various perturbing acceleration sources on an artificial earth-satellite.

Gravity Perturbations

Earth gravity perturbations of the satellite motion provide information on the earth's geoidal shape. It can be shown that the even zonal harmonics of the geopotential (the second, fourth, sixth and so on) produce secular perturbations in ω, Ω and M, and short-period perturbations in all six Keplerian elements, while the odd zonal harmonics produce long-period perturbations in all elements except the semi-major axis.

By far the largest departure of the earth from a spherical form is due to the flattening, or equatorial bulge. Earlier on, we saw that this departure can be represented by the second zonal harmonic, which is often given the symbol J_2. The corresponding orbital perturbations are:

$$\dot{a} = \dot{e} = \dot{i} = 0$$

$$\dot{\omega} = \frac{3n_0}{4(1-e^2)^2} \left[\frac{R_\oplus}{a}\right]^2 J_2 \ (1-5\cos^2 i)$$

$$\dot{\Omega} = \frac{3n_0}{2(1-e^2)^2} \left[\frac{R_\oplus}{a}\right]^2 J_2 \cos i$$

$$n = \dot{M} = n_0 - \frac{3n_0}{4(1-e^2)^{3/2}} \left[\frac{R_\oplus}{a}\right]^2 J_2 \ (3\cos^2 i - 1)$$

where n_0 ($= 2\pi/T$) is the mean motion for the two body-problem.

We see that the elements a, e, i undergo no secular change. In contrast, ω and Ω are subject to secular change and the mean motion is modified from what it would be for the two-body problem.
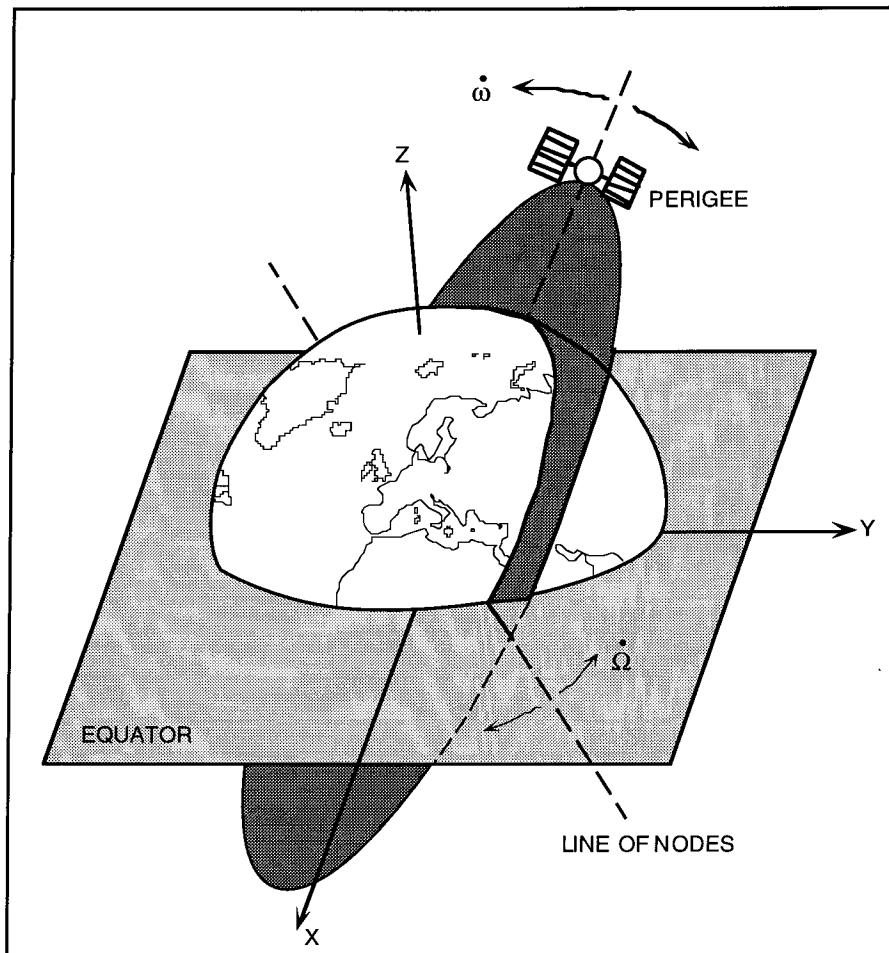
Figure 4.6-4

Nodal regression (Ω̇) and rotation of perigee (ω̇).

Basically, the gravitational pull of the equatorial bulge makes the satellite's orbit, (1) rotate around the earth's axis in a direction opposite to the satellite's motion, and (2) rotate around a line normal to the orbit plane in the direction of satellite motion. The first effect is known as *nodal regression* and the second as *rotation of perigee* (Figure 4.6-4). Thus, for example, an observer in space looking at the satellite orbiting the earth would see the orbit of an eastward-moving satellite gradually drifting to the west. The more flattened the earth, the faster the drift. For typical near-earth satellites $a = 1.1R_\oplus$ and $e \approx 0$ and using $GM_\oplus = 3.986\ 008 \times 10^5\ km^3s^{-2}$, $R_\oplus = 6\ 378.155$ km and $J_2 = 1.08 \times 10^{-3}$, we get:

$$\dot{\omega} \approx 3.6\ [5\cos^2 i - 1]\ {}^\circ day^{-1}$$

$$\dot{\Omega} \approx -6.7\ \cos i\ {}^\circ day^{-1}$$

$$n = \dot{M} \approx n_0\ [1 + 6.8 \times 10^{-4}(3\cos^2 i - 1)]$$

Since the actual drifts are quite rapid, we can measure their numerical values very accurately by allowing them to build up for several weeks or even months. All the even harmonics have similar a effect. Therefore, the values of the even harmonics are found by measuring the orbital drift of a number of satellites at different inclinations. The successive harmonics have different effects at different inclinations, and so they can be sorted out from one another. The odd harmonics give rise to different shapes in the northern and southern hemispheres and so to gravitational pulls that differ in the two hemispheres. The main effect of the unequal pulls is to make a satellite approach closer to the earth's centre when it is at its perigee. Again by measuring several orbits at different inclinations we can find the value of these harmonics.

Example 4.6-1

Assuming that $R_\oplus = 6.3780 \times 10^6$ m, $J_2 = 1.08 \times 10^{-3}$ and $i = 109°.8$ in Example 4.4-1 *determine* the node rate, perigee rotation rate and modified mean motion.

Solution

For convenience, we first convert the *mean motion*, n , from rad. s^{-1} to $° day^{-1}$:

$$n_0 = 4.65 \times 10^{-4} \text{ rad. } s^{-1} = 4.65 \times 10^{-4} \times (180/\pi) \times 86400 = 2301°.9 \text{ day}^{-1}$$

The *rotation rate of perigee* is:

$$\dot{\omega} = \frac{3n_0}{4(1-e^2)^2} \left[\frac{R_\oplus}{a} \right]^2 J_2 (1-5\cos^2 i)$$

$$= \frac{3 \times 2301.9}{4[1-(0.00444)^2]^2} \times \left[\frac{6.378}{12.27} \right]^2 \times 1.08 \times 10^{-3} \times (1-5\cos^2 109°.8)$$

$$= 0°.215 \text{ day}^{-1}$$

Similarly for the *node rate*, we have:

$$\dot{\Omega} = \frac{3n_0}{2(1-e^2)^2} \left[\frac{R_\oplus}{a} \right]^2 J_2 \cos i$$

$$= \frac{3 \times 2301.9}{2[1-(0.00444)^2]^2} \times \left[\frac{6.378}{12.27} \right]^2 \times 1.08 \times 10^{-3} \times \cos 109°.8$$

$$= 0°.341 \text{ day}^{-1}$$

The *modified mean motion* is:

$$n = \dot{M} = n_o - \frac{3n_o}{4(1-e^2)^{3/2}} \times \left[\frac{R_\oplus}{a}\right]^2 J_2 (3\cos^2 i - 1)$$

$$= 2301°9 - \frac{3 \times 2301.9}{4[1-(0.00444)^2]^{3/2}} \times \left[\frac{6.378}{12.27}\right]^2 \times 1.08 \times 10^{-3} \times (3\cos^2 109°8 - 1)$$

$$= 2301°6 \text{ day}^{-1}$$

Example 4.6-2

Calculate the J_2 for Mars *given* the following data for satellite Phobos: period = 0.32 days, semi-major axis = 9000 km, inclination = $1°1$, rate of regression of node = $0°56$ day^{-1} and equatorial radius of Mars = 3375 km.

Solution

The *mean motion* is:

$$n = 2\pi/P = \frac{2\pi}{0.32} = 19.6 \text{ rad day}^{-1}$$

Now:

$$J_2 = 2(1-e^2)^2 \left[\frac{a}{R_\oplus}\right]^2 \frac{1}{3n\cos i} \dot{\Omega}$$

Assuming a circular orbit (e=0), gives $J_2 = 2 \times 10^{-3}$.

Quite often a situation arises which dictates that a satellite be inserted into a specific orbit for the purpose of fulfilling a special mission objective. In general, the four orbital parameters a, e, i and Ω must be specified to define the shape of the orbit and the orientation of the orbital plane in space.

Example 4.6-3

What values must we assign to a, e, i and Ω to produce an orbital trajectory that is always over the same point on the surface of the earth (*geostationary orbit*)?

Solution

Since the earth rotates at a nearly uniform rate, we immediately conclude that the satellite must also have a uniform orbital motion. Kepler's second law tells us that this can only happen when the orbit is circular ie. e=0. Furthermore, to satisfy the geostationary criteria, $\dot{\Omega} = 0$, which implies that i = 0 (equatorial orbit), and that therefore, Ω is undefinable. Now only the semi-major axis remains to be defined.

The rotational period of the earth is about 86160 s and consequently, $n^2a^3 = GM_\oplus = 3.986 \times 10^{14}$ m^{3}s^{-2} gives:

$$a = \sqrt[3]{\frac{GM_\oplus}{n^2}} = \sqrt[3]{\frac{GM_\oplus}{(\frac{2\pi}{P})^2}} = \sqrt[3]{\frac{3.986 \times 10^{14} \text{ m}^3\text{s}^{-2}}{(\frac{6.283185}{86160 \text{ s}})^2}} = 4.21554 \times 10^7 \text{ m}$$

Lunar and Solar Accelerations

The primary effect of the moon and sun are to produce long-period perturbations of all Keplerian elements except the semi-major axis. Usually, these effects are quite small (<100 m) over short periods of time (several days) though exceptions to this rule can occur for high-eccentricity orbits (e>0.5). Here lunar perturbations can cause large oscillations in the eccentricity.

Drag Effects

The primary effect of atmospheric drag on a satellite is to produce secular decreases in the eccentrity until the satellite orbit is nearly circular, and then the semi-major axis will begin to decrease secularly. Atmospheric drag is usually the main factor in ending a satellite's life.

Solar Radiation Pressure

The perturbations due to solar radiation pressure are also small for satellites of normal construction, but can be large for balloon-type satellites or satellites powered by large solar panels (eg. the GPS satellites) since these have a large area-to-mass ratio. The effects are usually periodic, and if account is taken of the earth's shadow, all orbital elements are affected. However, if the satellite is in permanent sunlight, the semi-major axis remains unchanged.

REFERENCES

BATE, R.R., MUELLER, D.D. & WHITE, J., 1971. *Fundamentals of Astrodynamics*, Dover, New York, 455pp.

COOK, A.H., 1969. *Gravity and the Earth*, Wykeham, London, 95pp.

KING-HELE, D., 1967. The shape of the earth, *Scientific American*, **217**, 67-76.

LANGLEY, R., 1991. The orbits of GPS satellites, *GPS World*, March, 50-53.

SEEBER, G., 1993. *Satellite Geodesy*, Walter de Gruyter, Berlin, 531pp.

TORGE, W., 1991. *Geodesy*, Walter de Gruyter, Berlin, 264pp.

YUNCK, T.P., 1992. A new chapter in precise orbit determination, *GPS World*, October, 56-61.

EXERCISES

4.1   Which Keplerian elements define the shape of the orbit?

4.2   Which Keplerian elements define the orientation of the satellite orbit in inertial space?

4.3   Which Keplerian elements may used to locate the satellite in its orbit?

4.4   Distinguish between the true and eccentric anomalies.

4.5   Give Kepler's equation. What is it used for ?

4.6   Define the terms right ascension of ascending node and argument of perigee.

4.7   Starting from Newton's law of gravity and second law of motion show that the equations of motion of an earth-satellite relative to the earth are of the form:

$$\ddot{r} + \frac{GM_\oplus}{r^3} r = 0$$

Define all symbols used.

4.8   What simplifying assumptions are made in deriving the above equations.

4.9   Define the term mean motion.

4.10   What is a satellite state-vector?

4.11   Give the main perturbing forces acting on a near-earth satellite.

4.12   What are osculating orbital elements?

4.13   Give the equations of perturbed satellite motion in general terms.

4.14   The accelerations produced by the central part of the earth's gravity
       potential exceeds all other perturbing accelerations by about three
       orders of magnitude. Explain Why?

4.15   The earth's oblateness produces what two main perturbing effects on the
       motion of an earth-satellite? Hence explain how the earth's shape can be
       determined from satellite observations.

4.16   Describe the general effects of the earth's nonsphericity, the attraction of
       the sun and moon, atmospheric drag and solar radiation pressure on the
       orbit of a near-earth satellite.

4.17   Distinguish between a direct and retrograde orbit.

4.18   A satellite passed perigee at 66^d 15^h 51^m $51\overset{s}{.}809$ in 1981. Its eccentricity
       is e = 0.7248470 and its semi-major axis is a = 26 561.955 km. Compute
       n, M, E, v and r at midnight on day 67 of 1981. Take $GM_\oplus$ =
       3.986005×10^5 km^3/sec^2.

Answer:

n = 12.600640 rad/day, M = 244.73385 deg, E = 218.74286 deg, ν = 195.98736 deg, r = 41 578.848 km.

5 Coordinates and Coordinate Systems

5.1 INTRODUCTION

Geodetic positions are given in terms of coordinates which are simply a set of directly observed or calculated numbers which locate the point of interest in space. In geodesy we need four coordinates, three to define position relative to a set of reference axes attached to the earth in some prescribed manner and, the fourth to mark the time at which the coordinates were determined. We must also specify the time or epoch at which the reference axes were defined. The fourth dimension, time, is required because, with the accuracies now achievable (a few centimetres or better over 100 km), the sites which in practice define the reference coordinate system and the points whose positions we wish to determine in this reference system, must be considered to be in relative motion. Here we give the various coordinate types which occur in geodesy. We also define the two most important reference coordinate systems which are used and show how coordinates expressed in the one system may be transformed to coordinates in the other.
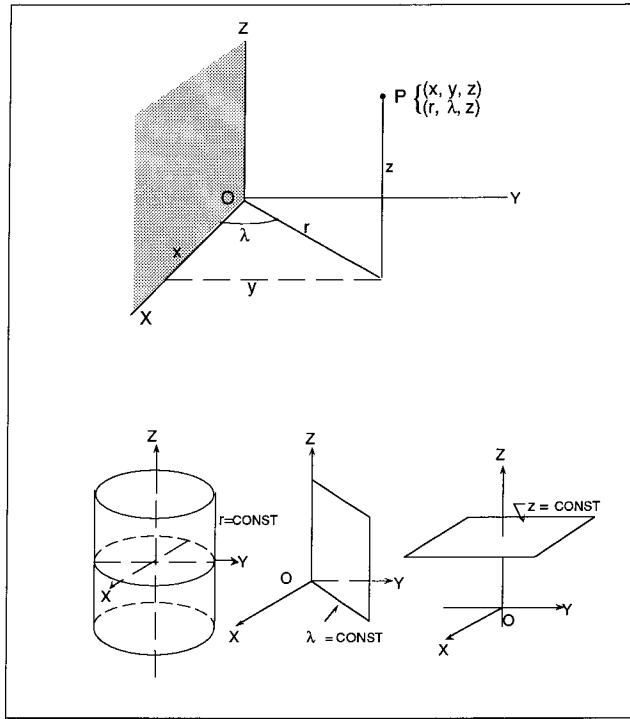
Figure 5.2-1

Cylindrical coordinates (r, λ, z).
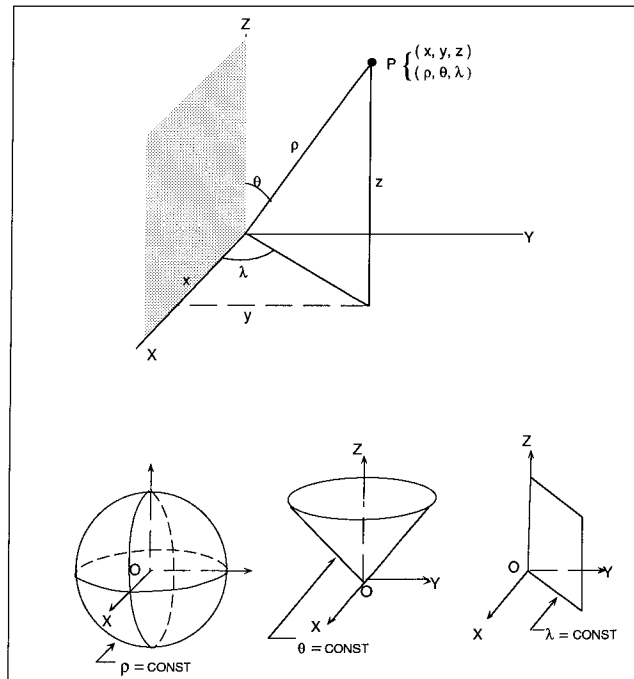
Figure 5.2-2

Spherical coordinates (ρ, θ, λ).

5.2   COORDINATE TYPES

The positions of points on the earth's surface are usually given either in *Cartesian*, *ellipsoidal* (geodetic) or *natural* (geographic) coordinates. Spherical and cylindrical coordinates are also sometimes used, particularly with space techniques. Most readers will be familiar with Cartesian coordinates, so we discuss only the other coordinate types.

Cylindrical Coordinates

A point P can be located by cylindrical coordinates (r, λ, z), Figure 5.2-1, as well as by rectangular Cartesian coordinates (x, y, z). Essentially, these are just the polar coordinates (r, λ) used instead of (x, y) in the plane $z = 0$ coupled with the z-coordinate. The transformation between Cartesian and cylindrical coordinates is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\lambda \\ r\sin\lambda \\ z \end{pmatrix} \tag{5.2-1}$$

and,

$$\begin{pmatrix} r \\ \lambda \\ z \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2} \\ \tan^{-1}\dfrac{y}{x} \\ z \end{pmatrix} \tag{5.2-2}$$

A *coordinate surface* is a surface along which one of the three coordinates is constant. If we hold r = const. and let λ and z vary, the locus of $P(r, \lambda, z)$ is then a right circular cylinder of radius r and axis along OZ. The locus λ = const. is a plane containing the Z-axis and making an angle λ with the XZ-plane.

Spherical Coordinates

A point P can be located by spherical coordinates (ρ, θ, λ), Figure 5.2-2. The transformation between Cartesian and spherical coordinates is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \rho\sin\theta\ \cos\lambda \\ \rho\sin\theta\ \sin\lambda \\ \rho\cos\theta \end{pmatrix} \tag{5.2-3}$$

and,

$$\begin{pmatrix} \rho \\ \theta \\ \lambda \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2+z^2} \\ \cos^{-1}\dfrac{z}{\sqrt{x^2+y^2+z^2}} \\ \tan^{-1}\dfrac{y}{x} \end{pmatrix}$$

(5.2-4)

The locus of points ρ = const. is the surface of a sphere of radius ρ with centre at O. The second spherical coordinate, θ, is the angle measured down from the Z-axis to the line OP. The locus of points θ = const. is a cone with vertex at O, axis OZ and generating angle θ. The third spherical coordinate, λ, is the same as the angle λ in cylindrical coordinates, namely, the angle from the XZ-plane to the plane through P and the Z-axis. Every point in the whole space can be given spherical coordinates restricted to the ranges:

$$\rho \ge 0, \, 0 \le \theta \le \pi , \, 0 \le \lambda < 2\pi$$

Because of the analogy between the surface of a sphere and the earth's surface, the Z-axis is sometimes called the polar axis, while θ is referred to as *co-latitude* and λ as *longitude*. One also speaks of meridians, parallels, and northern and southern hemispheres.

Example 5.2-1

The Cartesian coordinates of an observer on earth are:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -4\,648\,506.627 \\ 2\,546\,493.215 \\ -3\,536\,146.592 \end{pmatrix}, \text{m}$$

Calculate the corresponding cylindrical and spherical coordinates.

Solution

Cylindrical Coordinates:

$$\begin{pmatrix} r \\ \lambda \\ z \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2} \\ \tan^{-1}\dfrac{y}{x} \\ z \end{pmatrix} = \begin{pmatrix} 5\,300\,305.798 \\ 151^\circ\,17'\,8{.}''451 \\ -3\,536\,146.592 \end{pmatrix}$$

where r and z are in metres.
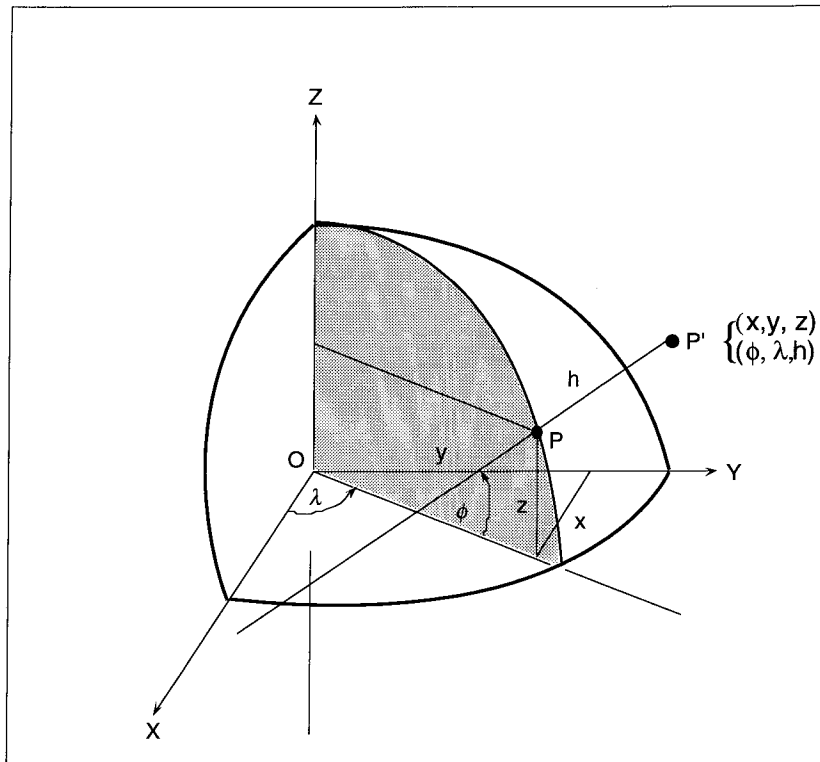
Figure 5.2-3

Geodetic or ellipsoidal coordinates (ϕ, λ, h).

Spherical Coordinates:

$$\begin{pmatrix} \rho \\ \theta \\ \lambda \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2+z^2} \\ \cos^{-1}\dfrac{z}{\sqrt{x^2+y^2+z^2}} \\ \tan^{-1}\dfrac{y}{x} \end{pmatrix} = \begin{pmatrix} 6\ 371\ 622.577 \\ 123°\ 42'\ 34.''528 \\ 151°\ 17'\ 8.''451 \end{pmatrix}$$

where ρ is in metres.

Ellipsoidal or Geodetic Coordinates

A point P can be located by ellipsoidal cordinates (ϕ, λ, h), Figure 5.2-3. For the meridian ellipse, it can be shown that (see below):

$$\begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} \nu\cos\phi \\ \nu(1-e^2)\sin\phi \end{pmatrix} \tag{5.2-5}$$

where,

$$v = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$ (5.2-6)

is the *prime vertical* radius of curvature at P and, a and e are respectively the semi-major axis and eccentricity of the meridian ellipse. For a point on the ellipsoid at longitude λ, we have:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} v\cos\phi \cos\lambda \\ v\cos\phi \sin\lambda \\ v(1-e^2)\sin\phi \end{pmatrix}$$ (5.2-7)

The point P is usually at a distance h, measured along the ellipsoidal normal, above or below the ellipsoid and the transformation between Cartesian and ellipsoidal coordinates, and vice versa is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (v+h)\cos\phi \cos\lambda \\ (v+h)\cos\phi \sin\lambda \\ [v(1-e^2)+h]\sin\phi \end{pmatrix}$$ (5.2-8)

$$\begin{pmatrix} \lambda \\ \phi \\ h \end{pmatrix} = \begin{pmatrix} \tan^{-1}\frac{y}{x} \\ \tan^{-1}\frac{z(v+h)\sin\lambda}{y[v(1-e^2)+h]} \\ \frac{x}{\cos\phi \cos\lambda} - v \end{pmatrix}$$ (5.2-9)

Equation (5.2-9) gives λ directly while, to obtain ϕ and h, it is necessary to iterate between the expressions for ϕ and h, since h is a function of ϕ and vice versa. We return to this problem in Section 6.5.

In geodesy, we use a base figure called the *reference ellipsoid* which approximates to the shape of the earth's figure over the globe or a particular region of interest. The meridians of this ellipsoid are ellipses, the shape of which are defined by a and e, while the parallels are circles of radius $v\sin\phi$. The coordinate surface h = const. is thus an ellipsoid which is concentric with the reference ellipsoid. The coordinate, ϕ, which is called the *geodetic latitude*, is the angle measured up or down from the XY-plane to the normal to the ellipsoid and passing through P. The locus of points ϕ = const. is a cone with vertex on the axis OZ, and generating angle 90°-ϕ. The coordinate, λ, the *geodetic longitude*, is the same as the angle λ in spherical and cylindrical coordinates, namely, the angle from the XZ-plane to the plane through P and the Z-axis. Every point in the whole space can be given geodetic coordinates restricted to the ranges:
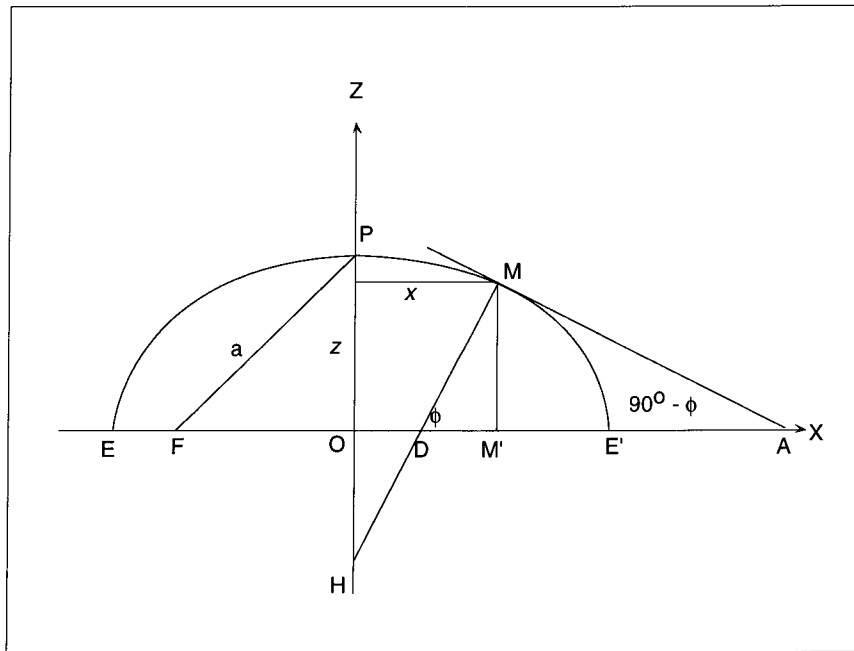
Figure 5.2-4

Geometry of meridian ellipse. Polar axis of ellipsoid (OP), equatorial diameter (EE'), tangent to ellipse at M (MA), normal to ellipse at M (MH), focus of ellipse (F), normal terminating in major axis (MD), semi-major axis (PF=a), semi-minor axis (OP=b), geodetic latitude (ϕ).

$(v+h) \geq 0, \; 0 \leq \phi \leq \pi/2 \, , \; 0 \leq \lambda < 2\pi$

Example 5.2-2

For the meridian ellipse, *show* that:

$$\begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} v\cos\phi \\ v(1-e^2)\sin\phi \end{pmatrix}$$

where,

$$v = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$

Solution

The general equation of an ellipse is given by (Figure 5.2-4):

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

where a and b are the semi-axes. Thus:

$$z^2 = b^2 - \frac{b^2}{a^2} x^2$$

Differentiating:

$$2z \, dz = -2 \frac{b^2}{a^2} x \, dx$$

or,

$$\frac{z}{x} = -\frac{b^2}{a^2} \frac{dx}{dz}$$

Since $\frac{dz}{dx}$ represents the tangent of the angle a tangent at any point M on the ellipse makes with the x-axis:

$$-\frac{dz}{dx} = \tan(90^\circ - \phi) = -\cot\phi$$

Hence:

$$\frac{z}{x} = \frac{b^2}{a^2} \tan\phi \tag{1}$$

By definition (Figure 5.2-4):

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

Thus:

$$\frac{z}{x} = \frac{b^2}{a^2} \tan\phi = (1 - e^2) \tan\phi \tag{2}$$

Also:

$$x^2 + \frac{z^2}{b^2} a^2 = a^2$$

and,

$$x^2 + \frac{z^2}{(1-e^2)} = a^2$$

or,

$$z^2 = (a^2-x^2)(1-e^2) \qquad (3)$$

Squaring (2) and comparing this with (3) gives:

$$z^2 = (1-e^2)^2 \tan^2\phi \, x^2 = (a^2-x^2)(1-e^2)$$

So:

$$a^2-x^2 = (1-e^2) \tan^2\phi \, x^2$$

or,

$$x^2[(1-e^2) \tan^2\phi +1] = a^2$$

Furthermore:

$$(1-e^2) \tan^2\phi +1 = \tan^2\phi -e^2\tan^2\phi + 1 = -e^2 \tan^2\phi + \sec^2\phi = \frac{1-e^2\sin^2\phi}{\cos^2\phi}$$

Hence:

$$x = \frac{a\cos\phi}{\sqrt{1-e^2\sin^2\phi}}$$

Similarly, eliminating x from (1) and using the equation of the ellipse:

$$x^2 = \frac{b^2-z^2}{1-e^2} = z^2/(1-e^2)^2\tan^2\phi$$

which reduces to:

$$z = \frac{a(1-e^2)\sin\phi}{\sqrt{1-e^2\sin^2\phi}}$$

Accordingly:

$$\begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} v\cos\phi \\ v(1-e^2)\sin\phi \end{pmatrix}$$

Example 5.2-3

If the *prime vertical curvature* of the ellipsoid at M is the length of the normal terminating in the minor axis of the meridian ellipse, *show* that:

$$\nu = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$

Solution

From geometry we see that:

$$\nu = \frac{x}{\cos\phi} = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$

Example 5.2-4

The geodetic coordinates of an observer on earth are:

$$\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} -25^\circ\ 54'\ 11.''0780 \\ 134^\circ\ 32'\ 46.''4570 \\ 397.5\ m \end{pmatrix}$$

Given that a = 6 378 160 m and e^2 = 0.006 694 541 855, *calculate* the corresponding Cartesian coordinates.

Solution

Now:

$$\nu = \frac{a}{\sqrt{1-e^2\sin^2\phi}} = 6\ 381\ 506.340\ m$$

Thus:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (\nu+h)\cos\phi\ \cos\lambda \\ (\nu+h)\cos\phi\ \sin\lambda \\ [\nu(1-e^2)+h]\sin\phi \end{pmatrix} = \begin{pmatrix} -3\ 026\ 602.098 \\ 5\ 018\ 916.297 \\ -2\ 508\ 732.689 \end{pmatrix}, m$$
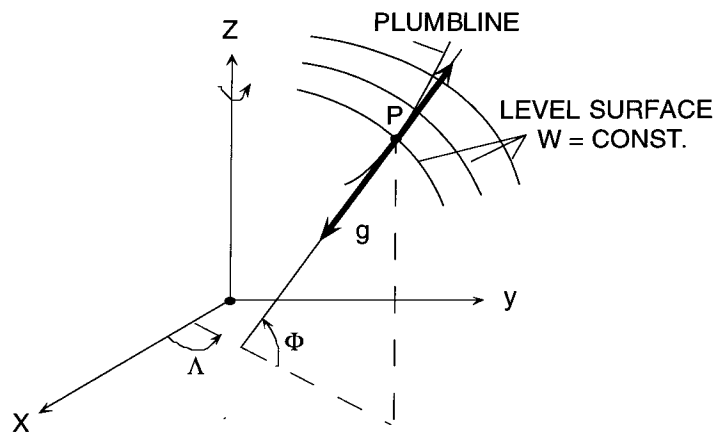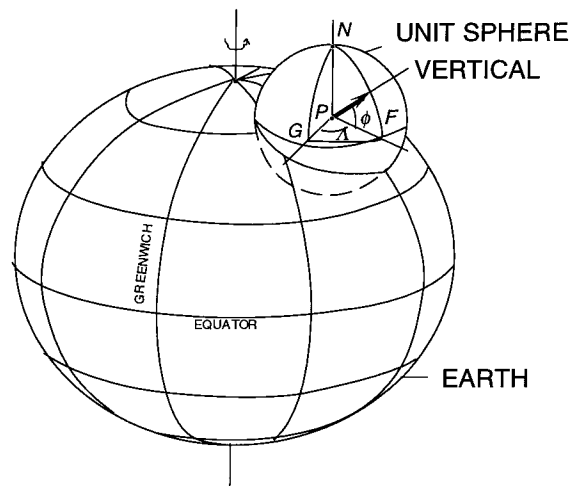
Figure 5.2-5

Geographical or natural coordinates (Λ, Φ, W). Line parallel to rotation axis (PN), plane parallel to equatorial plane and normal to rotation axis (GPF), meridian plane of P (NPF), plane parallel to Greenwich meridian plane (NPG).

Natural or Geographic Coordinates

The system of level surfaces and plumblines may be used as a three-dimensional curvilinear coordinate system. These coordinates may be observed directly, as opposed to rectangular Cartesian, cylindrical, spherical or geodetic coordinates which must be calculated from other measurements. The direction of the earth's spin axis and the position of the equatorial plane are well defined astronomically. The *geographic or astronomic latitude*, Φ, of a point P is the angle between the vertical at P and the equatorial plane (Figure 5.2-5); it is positive from the equator northward, negative to the south. Now consider a straight line through P parallel to the earth's spin axis. This line and the vertical at P together define the meridian plane of P. The angle between this meridian plane and the meridian plane of Greenwich is the *geographical or astronomical longitude*, Λ, of P; it is positive towards the east.

The geographical coordinates, latitude, Φ, and longitude, Λ, form two of the three spatial coordinates of P. As the third coordinate we take the distance between the geoid and P as measured along the curved plumbline. This is called the *orthometric height*, H (see also Section 6.3). Alternatively, we may take the geopotential, W, at P as the third coordinate. The transformation between (x, y, z) and (Λ, Φ, W) or (Λ, Φ, H) is of no interest here since this requires that W or H must be known as functions of x, y and z, which is generally not the case. The coordinate surfaces in natural coordinates are equipotential surfaces (W or H = const.), cones (Φ = const.), and planes (Λ = const.). Every point in the whole space can be given natural coordinates restricted to the ranges:

$$W \geq 0, 0 \leq \Phi \leq \pi/2 , 0 \leq \Lambda < 2\pi$$

H is generally positive; however, for points below sealevel H is negative.

5.3   REFERENCE COORDINATE SYSTEM TYPES

A coordinate system is defined by specifying the position of the *origin*, the orientation of the *fundamental plane* (the XY-plane),the *principal direction* (the X-axis), and the direction of the Z-axis. However, since the Z-axis is perpendicular to the fundamental plane it is only necessary to specify which direction is positive. The Y-axis is chosen so as to form a right-handed set of axes.

Coordinate systems are usually distinguished by the location of the origin. Examples of origins are: topocentric, geocentric, selenocentric, heliocentric, and barycentric. In topocentric coordinates, the origin is the location of the observer;
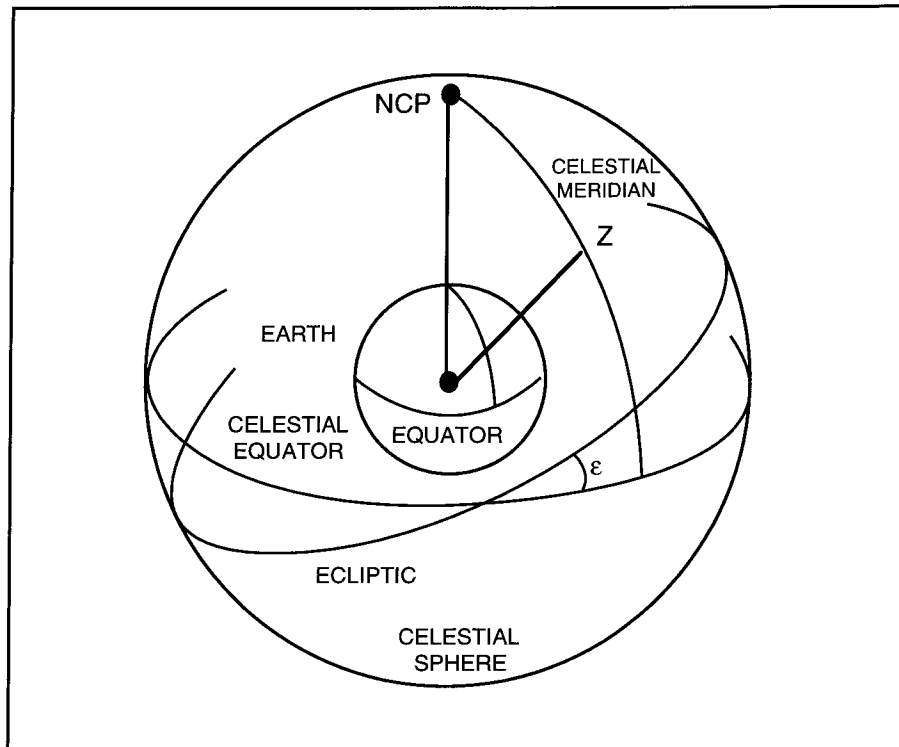
Figure 5.3-1

Celestial sphere showing ecliptic, celestial equator and celestial meridian.
North celestial pole (NCP), astronomical zenith (Z), obliquity of the ecliptic (ε).
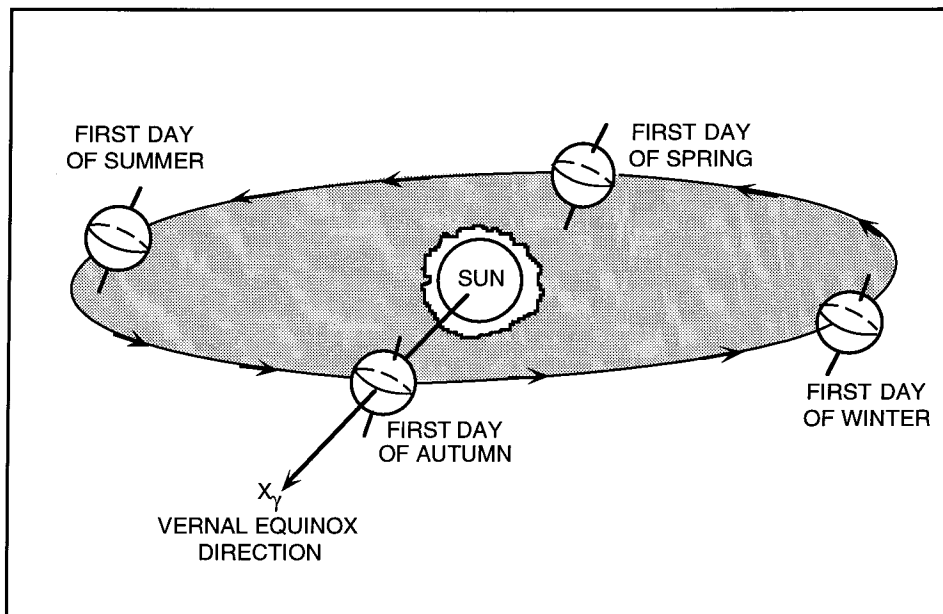
Figure 5.3-2

Ecliptic plane and vernal equinox. Seasons are for the northern hemisphere.

the selenocentre is the centre-of-mass of the moon and; barycentric coordinates refer to the centre-of-mass of the solar system.

The geocentre is usually taken as the origin of coordinates used in geodesy. This point is accessible indirectly through observations taken on artificial earth-satellites. Geodetic measurements provide both absolute and relative positions. Positions measured from the geocentre are called *absolute* or *point positions*. Satellite based positioning methods, such as LAGEOS laser ranging and GPS pseudoranging, are best suited for determining absolute positions. Conventional terrestrial surveys and some space-based geodetic methods, such as VLBI and differential GPS, give only positions relative to another site on the earth's surface. These *relative positions* must be combined with satellite methods based on ranging if absolute positions are desired.

Fundamental Planes

Frequently used fundamental planes are:

Equator. The equator is usually that of the earth. It is the plane normal to the axis of rotation, and the positive direction is that of the angular velocity vector of the rotation. The point where the earth's rotation axis pierces an imaginary sphere of infinite radius, centred at the geocentre, is called the *celestial pole* . This imaginary sphere is called the *celestial sphere* (Figure 5.3-1).

Orbital Plane. The plane of an orbit is defined by two-body motion and the positive direction is that of the angular momentum vector of the system.

Ecliptic. The ecliptic is the special case of the plane of an orbit. It is the plane of the earth's orbit about the sun (Figure 5.3-2). The positive direction is that of the earth-sun system's angular momentum vector. The point where the angular momentum vector pierces the celestial sphere is called the *ecliptic pole* (Figure 5.3-1).

Horizon Plane. The fundamental plane is the horizon. The positive direction is that of the local vertical.

Principal Directions

The principal direction is usually specified by giving the sense along the intersection of the fundamental plane with some other plane. The other plane may be a meridian plane, an equatorial plane, or another orbital plane. A meridian plane is a plane which contains the axis of rotation of one of the principal gravitating bodies. Frequently used principal directions are:

Greenwich Meridian or Prime Meridian. The Greenwich meridian is the earth's meridian plane that passes through the former Royal Observatory at Greenwich, United Kingdom.
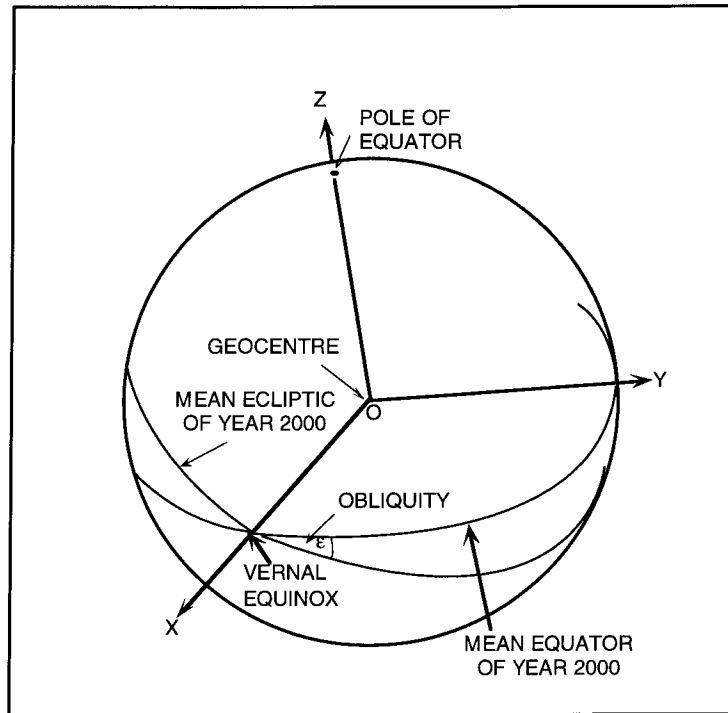
Figure 5.4-1
Geocentric-equatorial inertial (space-fixed) coordinate system. The barycentric equivalent of this system is called the international celestial reference frame (ICRF).
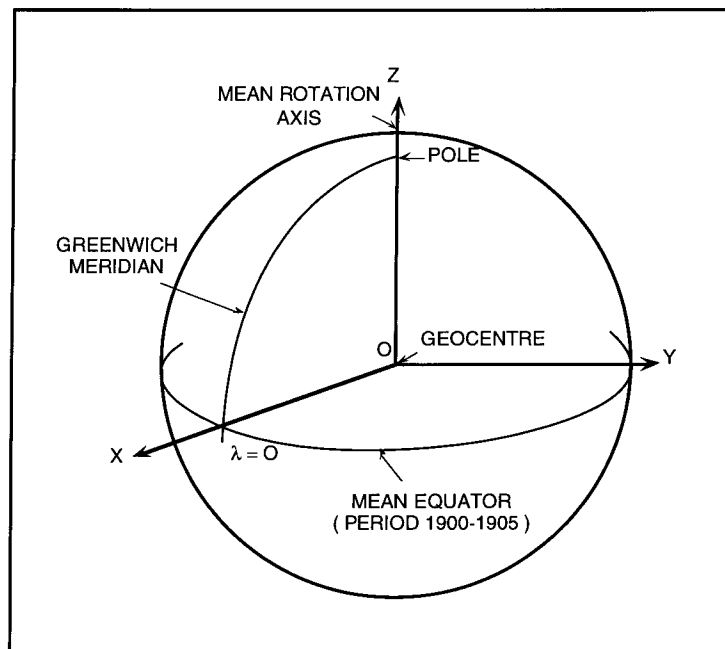
Figure 5.4-2
International terrestrial reference frame (ITRF).

Vernal Equinox or Equinox. The vernal equinox is the fundamental principal direction used in space geodesy and astronomy (Figure 5.3-2). It is defined as the intersection of the ecliptic and the earth's equator with the positive sense being from the earth to the sun as the sun crosses the equator from south to north. The point where the principal direction pierces the celestial sphere is also called the *first point of Aries* (Υ) because it points in the general direction of the constellation Aries.

Ascending Node. The ascending node is the intersection of an orbital plane and the fundamental plane with the positive sense being from the origin towards the orbiting body as it crosses the fundamental plane from south to north. Hence, Υ is an ascending node.

5.4   GEODETIC REFERENCE COORDINATE SYSTEMS AND DATUMS

Below we describe the main coordinate systems used in geodesy. We also briefly discuss the concept of a geodetic datum. The two most important geodetic reference coordinate systems are:

Geocentric-Equatorial Inertial (Space-Fixed) System

The geocentric-equatorial inertial $(X_\Upsilon, Y_\Upsilon, Z_\Upsilon)$ coordinate system (Figure 5.4-1) has its origin at the geocentre; the $X_\Upsilon Y_\Upsilon$-plane is the earth's mean equator at a particular epoch, namely 12h TDB on 1 January 2000, designated J2000.0, and; the X_Υ-axis is the intersection of the mean equator and mean ecliptic at epoch J2000.0 and points towards Υ. We leave the term <u>mean</u> undefined for the moment. The barycentric equivalent of this system is called the *International Celestial Reference Coordinate Frame* or ICRF. In practice, the directions of the $X_\Upsilon, Y_\Upsilon, Z_\Upsilon$-axes are defined by the observed positions of celestial radiosources.

Geocentric-Equatorial Rotating (Earth-Fixed) System

The geocentric earth-fixed $(X_\oplus, Y_\oplus, Z_\oplus)$ coordinate system (Figure 5.4-2), which is also called the *International Terrestrial Reference Coordinate Frame* or ITRF, has its origin at the geocentre; the $X_\oplus Y_\oplus$-plane is the mean equator of the period 1900-1905 and; the $X_\oplus$-axis lies in the plane through Greenwich and the mean rotation axis of the year 1984 as defined by the International Earth Rotation Service in Paris. In practice, the ITRF is defined by a global network of observatories and their relative motions.

Other reference coordinate systems which occur in geodesy, are: (1) the topocentric horizon system and, (2) the orbit plane or perifocal system. The
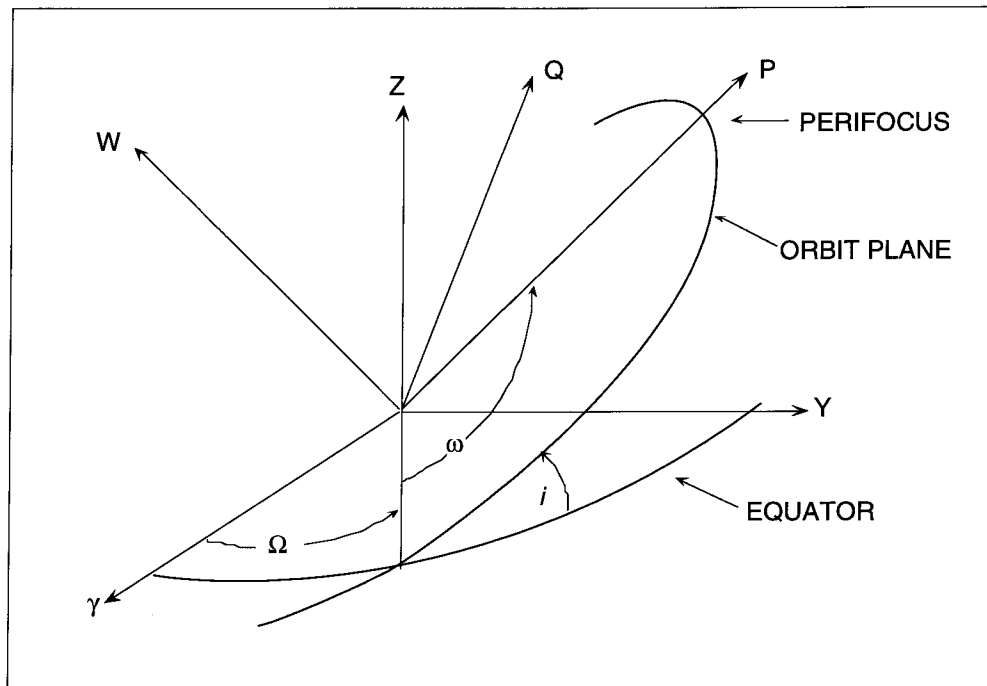
Figure 5.4-3
Orbit plane or perifocal coordinates.

topocentric horizon system and its geodetic counterpart, the geodetic horizon system, are discussed in Section 6.5. Here we deal only with orbit plane coordinates.

Orbit Plane or Perifocal Coordinate System

The orbit plane or perifocal (P, Q, W) coordinate system (Figure 5.4-3) has its origin at the centre-of-mass of the parent body eg. the geocentre in the case of an earth-satellite; the PQ-plane is the plane of the orbit; P points towards the perifocus; Q is in the orbit plane advanced to P by a right angle in the direction of increasing true anomaly, and; W completes the right-handed system.

Geodetic Datums

A geodetic datum is more than a reference coordinate system; it is a reference surface for horizontal and vertical position as well. In general, the horizontal and vertical reference surfaces are not the same. Most frequently, a horizontal datum is an ellipsoid of revolution while the commonly used vertical geodetic datum is the geoid. On each such datum a network of control stations is emplaced to provide the user with accurate position coordinates.

TABLE 5.4-1 *Relationships Between Selected Geodetic Datums and WGS84*

Datum	Ellipsoid Parameters a (m)	1/f	Datum Shifts (m) Dx	Dy	Dz	Coordinate System Rotations (") ε	ψ	ω	Scale (ppm) ΔS
WGS 84	6 378 137	298.257 223 563	0	0	0	0	0	0	0
North American 1983	6 378 137	298.257 222 101	0.42	0.95	-0.62	-0.012	-0.006	0.012	-0.1364
European 1950	6 378 388	297.0	-102	-102	-129	0.413	-0.184	0.385	2.4664
South American 1969	6 378 160	298.25	-56	-3	-38	0.123	-0.569	-0.158	-0.6412
Australian Geodetic 1966	6 378 160	298.25	-127	-50	153	0.058	-0.018	-0.089	1.2065
Indian	6 377 276.345	300.8017	227	803	274	-0.444	-0.645	-0.353	6.5931
Australian Geodetic 1994	6378160	298.25	-116.0	-50.5	141.7	-0.230	-0.390	-0.344	0.098

Eight parameters are needed to define a horizontal geodetic datum: two to determine the dimensions of the ellipsoid, three to locate its centre, and three to specify the orientation of the axes. Section 6.2 describes how this definition is accomplished by terrestrial geodetic methods. Over time, literally hundreds of such horizontal datums have been created. Generally, the ellipsoids were chosen so that they would fit as closely as possible the geoid for the region. The result is that the ellipsoids are not geocentric, that the centres do not coincide, that they may have different dimensions and that their respective axes are slightly askew to oneanother. Modern practice is to use satellite methods to define a datum. Invariably, these satellite datums are geocentric. Table 5.4-1 gives the dimensions of some of the horizontal datums in current use.

The coordinates initially computed within GPS receivers refer to the World Geodetic System 1984 (WGS84). This is a geocentric system defined by space geodetic methods. The axes of the ellipsoid are aligned with those of the ITRF, the semi-major axis is a = 6 378 137 m and, the flattening is 1/f = 298.257 223 563. Table 5.4-1 gives the relationships between selected geodetic datums and WGS84.

5.5 TRANSFORMATIONS BETWEEN COORDINATE SYSTEMS AND DATUMS

The components of a vector, such as the position of the oberver or of a satellite, may be expressed in any of the coordinate systems described above. So it often becomes necessary to transform these components from one system to another. We deal with this problem in this section. We note that a coordinate transformation merely converts the components of the vector in one coordinate system to components in another system - nothing else. The vector still has the same length and direction after the coordinate transformation, and it still represents the same thing.
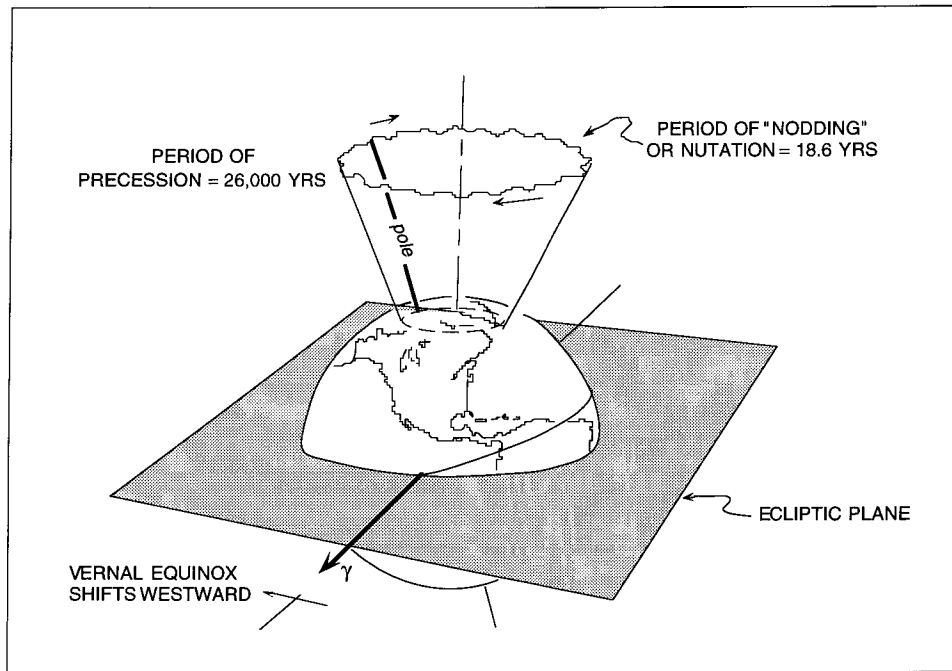
Figure 5.5-1
Precession and nutation.
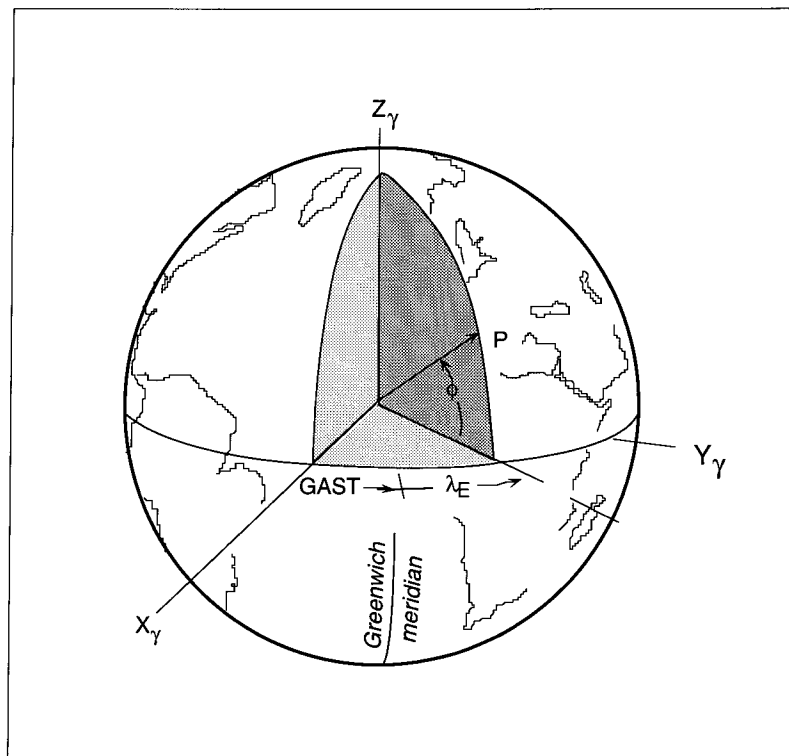
Figure 5.5-2
Greenwich apparent sidereal time (GAST).

Space-Fixed to Earth-Fixed Coordinates

The earth's pole of rotation is not fixed in space but precesses and nutates due principally to the torques exerted by the gravitational fields of the moon and sun on the equatorial bulge. *Precession* is the slow circular motion of the rotation pole with respect to inertial space with a period of about 26 000 years (Figure 5.5-1). *Nutation* is a more rapid motion, superimposed on the precession, and is comprised of a number of oscillations ranging in period from 14 days to 18.6 years (Figure 5.5-1). Both motions are predictable to a high degree of accuracy. Each produces a motion of the instantaneous equator and equinox with respect to the fixed equator and equinox of a given date. Precession changes the celestial longitude of a point on the earth by about 50 arcsec/year. Celestial longitude is longitude in the usual sense but with respect to the ITRF. Nutation changes both the celestial longitude and latitude by as much as 20 arcsec/18.6 years. Removing the effects of nutation produces a fictitious equator and equinox called the mean *equator-of-date* and mean *equinox-of-date*.

Two additional motions must be applied to relate the celestial and terrestrial reference coordinate systems. The first is a rotation about the celestial pole through the angle between the equinox and the Greenwich meridian (Figure 5.5-2). Although this is an angle and its units are radians or degrees it is called the *Greenwich Apparent Sidereal Time* (GAST). The second accounts for the fact that the position of the earth's equatorial plane relative to the crust shifts slightly with time. This is known as *polar motion*. Polar motion is too irregular to be predictable and, along with changes in the length-of-day (UT), is determined from observations. Graphs of polar motion show a secular trend, a strong periodic component, known as the *Chandler wobble*, with a period of about 14 months, and a component with an annual period. The amplitude of polar motion is about 0.1-0.3 arcsec (Figure 5.5-3).

Denoting the ICRF coordinates by the vector R_Υ and the corresponding ITRF coordinates by the vector $r_\oplus$, we have:

$$r_\oplus = [W] \, [T] \, [N] \, [P] \, R_\Upsilon \qquad\qquad (5.5\text{-}1)$$

where W, T, N and P are matrices which respectively account for polar motion, GAST, nutation and precession.

The transformation from ITRF to ICRF coordinates is:

$$R_\Upsilon = = ([W] \, [T] \, [N] \, [P])^{-1} \, r_\oplus = ([P] \, [N] \, [T] \, [W])^{T} \, r_\oplus \qquad\qquad (5.5\text{-}2)$$

where the superscript T denotes the transpose of the matrix.

Example 5.5-1

Neglecting precession, nutation and polar motion *show that:*
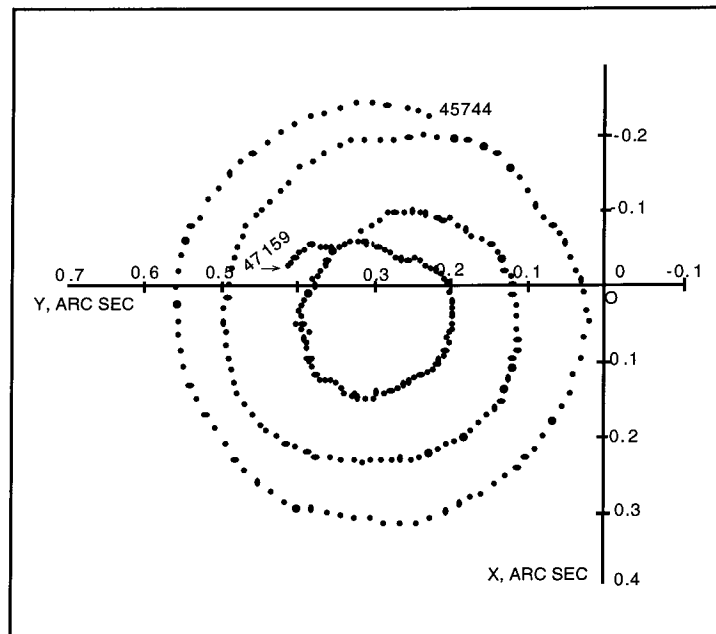
Figure 5.5-3

Path of earth's rotation pole from 14 Feb. 1984 (MJD 45744) to 30 Dec. 1987 (MJD 47159). OX is parallel to Greenwich meridian plane.

$$\begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix}$$

where θ = GAST.

Hence *calculate* $X_\Upsilon, Y_\Upsilon, Z_\Upsilon$ for a site at latitude $\phi = 30°N$, longitude $\lambda = 191°E$, $R_\oplus = 6\,371$ km and $\theta = 304°$.

Solution

Figure 5.5-4 shows the geometry of the problem. Since we may neglect precession, nutation and polar motion the Z_Υ and $Z_\oplus$-axes coincide and both point out of the plane of the paper. Clearly:

$$X_\Upsilon = X_\oplus \cos\theta - Y_\oplus \sin\theta$$

$$Y_\Upsilon = X_\oplus \sin\theta + Y_\oplus \cos\theta$$

$$Z_\Upsilon = Z_\oplus$$

that is,
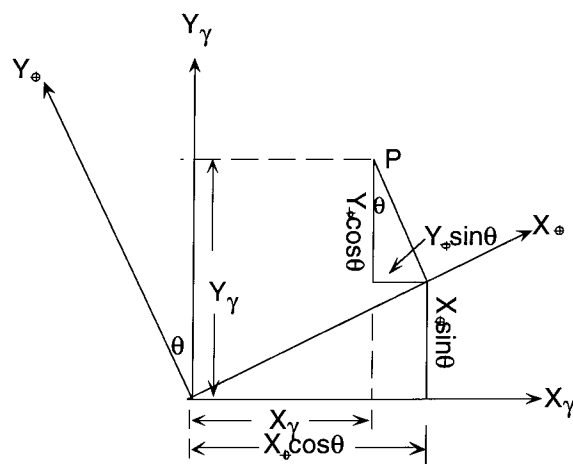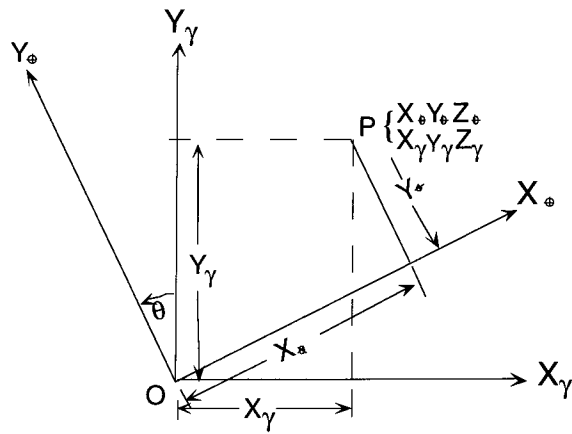
Figure 5.5-4
Relationship between earth-fixed and space-fixed coordinates neglecting precession and nutation.

$$\begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix}$$

Also:

$$\begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix} = \begin{pmatrix} R_\oplus \cos\phi\cos\lambda \\ R_\oplus \cos\phi\sin\lambda \\ R_\oplus \sin\phi \end{pmatrix} = \begin{pmatrix} -5\ 416\ 076.8 \\ -1\ 052\ 778.7 \\ 3\ 185\ 500.0 \end{pmatrix}, m$$

and,

$$\begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} = \begin{pmatrix} \text{-3 901 424.8} \\ \text{3 901 424.8} \\ \text{3 185 500.0} \end{pmatrix}, \text{m}$$

Orbit Plane to Space-Fixed Coordinates

Orbit plane coordinates transform to space-fixed coordinates according to the relation:

$$\begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} = \mathbf{R} \begin{pmatrix} P \\ Q \\ W \end{pmatrix} \qquad (5.5\text{-}3)$$

where,

$$\mathbf{R} = \begin{pmatrix} \cos\omega\cos\Omega\text{-}\sin\omega\cos i\sin\Omega & \text{-}\sin\omega\cos\Omega\text{-}\cos\omega\cos i\sin\Omega & \sin\Omega\sin i \\ \cos\omega\sin\Omega\text{+}\sin\omega\cos i\cos\Omega & \text{-}\sin\omega\sin\Omega\text{+}\cos\omega\cos\Omega\cos i & \text{-}\cos\Omega\sin i \\ \sin\omega\sin i & \cos\omega\sin i & \cos i \end{pmatrix}$$

The transformation from space-fixed to orbit plane coordinates is:

$$\begin{pmatrix} P \\ Q \\ W \end{pmatrix} = \mathbf{R}^\mathsf{T} \begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} \qquad (5.5\text{-}4)$$

Example 5.5-2

The orbit plane coordinates of a satellite are:

$$\begin{pmatrix} P \\ Q \\ W \end{pmatrix} = \begin{pmatrix} \text{-9.0160} \\ \text{-8.3811} \\ 0 \end{pmatrix} \times 10^6, \text{m}$$

Given that:

$$\begin{pmatrix} \omega \\ i \\ \Omega \end{pmatrix} = \begin{pmatrix} 261.75 \\ 109.84 \\ 29.57 \end{pmatrix} \text{deg.}$$

Find the space-fixed coordinates.
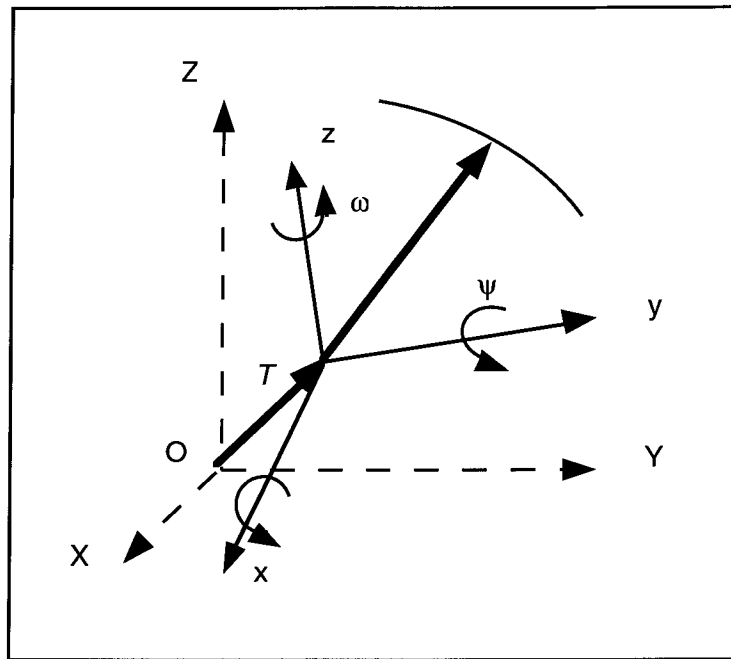
Figure 5.5-5
Differential transformation between Cartesian coordinate systems.

Solution

The components of the matrix **R** are:

$$R = \begin{pmatrix} -0.3569 & -0.8107 & 0.4642 \\ -0.5235 & -0.2380 & -0.8181 \\ 0.7737 & 0.5349 & -0.3394 \end{pmatrix}$$

Multiplication gives the space-fixed coordinates of the satellite ie.:

$$\begin{pmatrix} X_\text{T} \\ Y_\text{T} \\ Z_\text{T} \end{pmatrix} = \begin{pmatrix} -4.3934 \\ -6.4442 \\ 9.5241 \end{pmatrix} \times 10^6 \text{ metres}$$

Datum Transformations

Often we have coordinates on a global datum such as WGS84 but require them on another datum eg. AGD66. This involves what is commonly called a datum transformation which simply is a transformation of the coordinates from one datum to another. There are many ways of doing this. We give only the similarity transformation method here. This preserves the shape of figures, so angles will remain unaltered. However, the lengths of the sides and the positions of the corners may change.

Figure 5.5-5 shows the coordinate system **X**(X, Y, Z) which is related to the coordinate system **x**(x, y, z) by the translation vector **T** = $(\Delta X, \Delta Y, \Delta z)^T$ between the origins of the two coordinate systems and the small rotations $(\varepsilon, \psi, \omega)$ around the (x, y, z) axes. The transformation equation expressed in the **X** coordinate system is:

$$\mathbf{X} = (1+\Delta s) \ \mathbf{R} \ \mathbf{x} + \mathbf{T}$$ (5.5-5)

where $(1+\Delta s)$ denotes the scale factor between the systems and **R** is a rotation matrix.

Usually, the axes will be almost aligned, so that we may write:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = (1+\Delta s) \begin{pmatrix} 1 & \omega & -\psi \\ -\omega & 1 & \varepsilon \\ \psi & -\varepsilon & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta X \\ \Delta Y \end{pmatrix}$$ (5.5-6)

where the angles ε, ψ and ω are in radians.

Table 5.4-1 gives the transformation parameters between selected geodetic datums and WGS84. These parameters are obtained by comparing coordinates at common sites. Usually, there are many more points then there are parameters. Accordingly, the solution is overdetermined and, therefore, more precise. The datum shifts are the offsets from the centre of the WGS84 ellipsoid; the coordinate system rotations represent the misalignment of the regional datum coordinate system axes relative to those of WGS84, and; the scale parameters account for the difference between each datum's length scale and that of the WGS84 datum.

Example 5.5-3

The WGS84 coordinates of a point on the earth are:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -4\ 446\ 476.915 \\ 2\ 678\ 127.000 \\ -3\ 696\ 251.423 \end{pmatrix}, m$$

Transform these to coordinates on the AGD66. *Use* the values given in Table 5.4-1.

Solution

The values listed in Table 5.4-1 are for transforming from the local datum to WGS84. To transform the other way the signs must be reversed. So we have:

$$(1+\Delta s) = 0.999\ 993\ 407$$

$$R = \frac{1}{206264.8062} \begin{pmatrix} 1 & 0.353 & -0.645 \\ -0.353 & 1 & 0.444 \\ 0.645 & -0.444 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} -227 \\ -803 \\ -274 \end{pmatrix}, m$$

where the factor 206 264.8062 is the number of seconds in a radian.

This gives:

$$R\ x_{WGS84} = \begin{pmatrix} -4\ 446\ 460.774 \\ 2\ 678\ 126.654 \\ -3\ 696\ 271.092 \end{pmatrix}, m$$

$$(1+\Delta s)\ R\ x_{WGS84} = \begin{pmatrix} -4\ 446\ 431.458 \\ 2\ 678\ 108.997 \\ -3\ 696\ 246.722 \end{pmatrix}, m$$

$$X_{AGD66} = (1+\Delta s)\ R\ x_{WGS84} + T = \begin{pmatrix} -4\ 446\ 658.458 \\ 2\ 677\ 305.997 \\ -3\ 696\ 520.722 \end{pmatrix}, m$$

REFERENCES

BOMFORD, G., 1971. *Geodesy*, Oxford University Press, Oxford, 855pp.

ESCOBAL, P.D., 1985. *Methods of Orbit Determination*, R. E. Krieger Publ. Co., Malabar, Florida, 479pp.

HARVEY, B.R., 1994. *Practical Least Squares and Statistics for Surveyors*, 2nd Edition, Monograph 13, School of Geomatic Engineering, The University of NSW, Kensington, NSW, 319pp.

HEISKANEN, W.A. & MORITZ, H., 1967. *Physical Geodesy*, W.H. Freeman, San Francisco, 364pp.

KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A. & COLLINS, J., 1987. *Surveying with the Global Positioning System*, F. Dümmler, Bonn, 128pp.

MANNING, J. & STEED, J.B., 1994. Geodetic coordinate systems in Australia, *Preview*, **51**, 19-26.

STEED, J.B., 1990. A practical approach to transformation between commonly used reference systems, *The Australian Surveyor*, **35** (3), 248-2644.

STEED, J.B., 1995. The geocentric datum of Australia, *Surveying World*, **4** (1), 14-17.

THOMAS, G.B., 1960. *Calculus and Analytical Geometry* (3rd edition), Addison-Wesley, Reading, Mass., 1010pp.

WAHR, J., 1986. The earth's inconstant rotation, *Sky & Telescope*, June, 545-549.

EXERCISES

5.1   Distinguish between absolute and relative positioning.

5.2   Name a space geodetic technique which gives only relative positions.

5.3   Distinguish between geocentric, selenocentric, topocentric, heliocentric and barycentric coordinates.

5.4   What is a coordinate surface?

5.5   Name the coordinate surfaces in (a) Cartesian coordinates, (b) ellipsoidal coordinates, and (c) natural coordinates.

5.6   Why are ellipsoidal coordinates preferred in geodesy?

5.7   Using a diagram to illustrate your answer define the terms eccentricity and flattening of an ellipsoid of revolution.

5.8   Distinguish between geodetic and geocentric latitude. Hence show that the two are the same at four points on the earth and that they differ by about 11 arcmin. at $\phi = 45^{\circ}$.

5.9   Define the terms meridian and prime vertical radius of curvature.

5.10   What are natural coordinates? How do they differ from their geodetic counterparts?

5.11   What is a triaxial ellipsoid?

5.12   For the meridian ellipse show that:

$$f = 1 - \sqrt{1-e^2}$$

where,

f = flattening

e = eccentricity

Hence find e^2 for $\frac{1}{f} = 298.257$.

Answer:

$e^2 = 0.006\ 694385$

5.13   The geodetic coordinates of an observer on earth are:

$$\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} -35^{\circ}\ 38'\ 10\rlap{.}''5140 \\ 148^{\circ}\ 56'\ 21\rlap{.}''5252 \\ 1349.893\ m \end{pmatrix}$$

Given that a = 6 378 137.0 m and $\frac{1}{f} = 298.257$ calculate the corresponding Cartesian coordinates.

Answer:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -4446476.915 \\ 2678127.000 \\ -3696251.423 \end{pmatrix}, m$$

5.14   Give the basic components of a reference coordinate system.

5.15   Define the following: equator, ecliptic and horizon plane. Why are these used as fundamental planes?

5.16   What is the Vernal Equinox? Why is it called the first point of Aries?

5.17   Define the earth-fixed and space-fixed coordinate systems used in geodesy. Why do we need both?

5.18   How is the earth-fixed system determined in practice?

5.19   How is the space-fixed system determined in practice?

5.20   What is the relationship between the earth's spin axis and latitude?

5.21   Distinguish between precession and nutation. Give their principal periods and amplitudes.

5.22   Define Greenwich Apparent Sidereal Time.

5.23   What is polar motion? Give the three principal components.

5.24   Give the transformation from space-fixed to earth-fixed coordinates in a general way.

5.25   Neglecting precession, nutation and polar motion show that:

$$\begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix}$$

where θ = GAST.

5.26 What is the space-fixed position of a point 6.378 km above the reference ellipsoid, on the equator, at $302^\circ\!.704$ east longitude, for GAST = 8.6245 rad.? Take a = 6378.165 km and e = 0.08181.

Answer:

$$\begin{pmatrix} X_\Upsilon \\ Y_\Upsilon \\ Z_\Upsilon \end{pmatrix} = \begin{pmatrix} -0.697 \\ 0.718 \\ 0 \end{pmatrix}, \text{km}$$

5.27 The AGD84 coordinates of a point are:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -4\ 803\ 113.750 \\ 2\ 679\ 285.334 \\ -3\ 219\ 263.574 \end{pmatrix}, \text{m}$$

Transform these to WGS84 coordinates. *Use* the values listed in Table 5.4-1.

Answer:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} -4\ 803\ 240.776 \\ 2\ 679\ 230.706 \\ -3\ 219\ 110.130 \end{pmatrix}, \text{m}$$

6 Terrestrial Geodetic Methods

6.1 INTRODUCTION

The geodetic measurements which are made to determine positions on the earth's surface can be divided into four categories. They are: (1) astronomic determinations of latitude, longitude and azimuth; (2) gravity readings; (3) observed horizontal angles, distances, zenith angles, and height differences; and (4) measurements to artificial earth satellites, the moon, and extragalactic radiosources. The measurements in categories (1), (2) and (3) form the basis of the terrestrial geodetic positioning methods, while those in category (4) form the basis of the space-based geodetic methods. Here we briefly discuss the terrestrial methods.

The discussion of the terrestrial methods is usually divided into three parts: (i) the establishment of horizontal control; (ii) the establishment of vertical control; and (iii) the measurement of gravity. This subdivision is one of observational convenience, a consequence of the instruments used and of the geodetic objectives with which the measurements have been identified. These three aspects are nevertheless closely related. Consider, for example, the measurement of distance between two stations on the earth's surface. The reduction of the measurement to the reference ellipsoid (see below) requires that the height above the geoid and the geoid-ellipsoid separation be known at the two points. The height above the geoid is determined from levelling and gravity, while the geoid-ellipsoid separation is obtained from gravity.

MEASURED-
ALL ANGLES
IN EVERY
TRIANGLE.

LENGTHS AB
AND CD

TRIANGULATION NETWORK

OPEN TRAVERSE

CLOSED TRAVERSE

CLOSED TRAVERSE

TRAVERSES

MEASURED-
ALL SIDES IN
THE NETWORK

TRILATERATION NETWORK
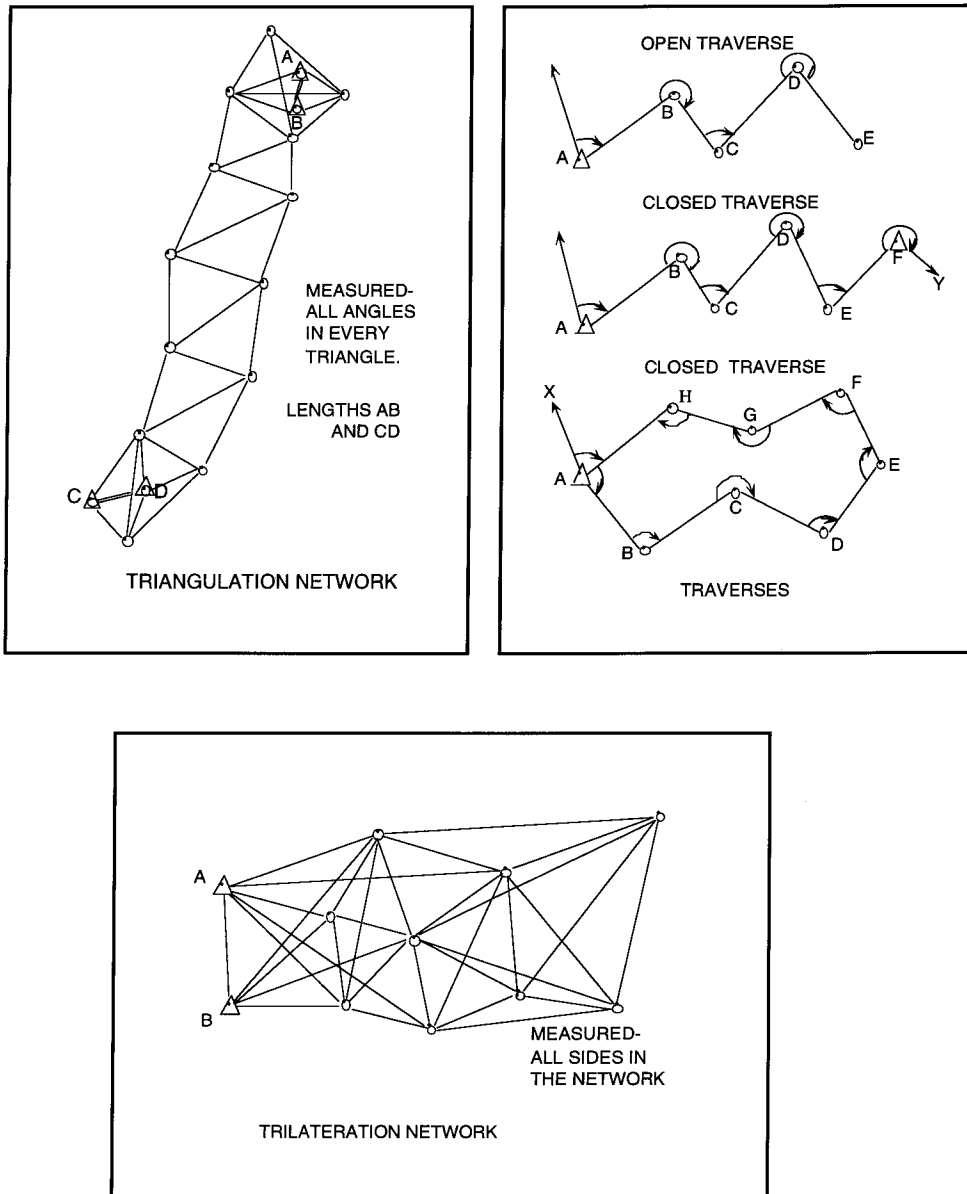
Figure 6.2-1

Terrestrial geodetic methods: triangulation, trilateration and traversing.

Alternatively, the heights may be determined from vertical angles (trigonometrical levelling), but only when the orientation of the observers vertical has been determined from either gravity surveys or from astronomical position observations.

6.2   HORIZONTAL CONTROL SURVEYS

Horizontal control networks have traditionally been established by a process of triangulation, trilateration, or traverse surveys (Figure 6.2-1). *Triangulation* is the process of determining the relative positions of a network of control points by theodolite measurements of directions or angles between intervisible stations. In *trilateration*, the relative positions are determined from distance measurements between sites using mainly optical ranging instruments. *Traverse surveys* consist of measurements of both directions and distances along a series of connected lines, rather than a network of stations.

The conventional computation process is to reduce all measured quantities to a reference ellipsoid. Two coordinates of each station are computed on this ellipsoid: geodetic latitude, ϕ, and geodetic longitude, λ. The third dimension, height, h, above the ellipsoid, is determined separately, either from levelling and gravity data (see below) or from vertical angles. Traditionally, the reference ellipsoid approximates the geoid over the survey area. A *datum point* is selected whose coordinates are its astronomical latitude, Φ, and longitude, Λ. All measured directions and distances are then projected along normals onto the ellipsoid and the geodetic latitude and longitude are computed relative to this reference surface. Astronomical azimuths determine the orientation of the networks with respect to the reference ellipsoid. Astronomical measurements made at points other than the datum point will not, in general, equal their geodetic counterparts computed from the triangulation chain, even when no errors have been incurred in the measurements and computations. The differences in latitude and longitude reflect the nonalignment of the normals to the ellipsoid with the observer's vertical. These are the deflections of the vertical defined in Chapter 2. Computation schemes for three-dimensional networks have also been explored but without much success, mainly due to atmospheric refraction effects which limit the accuracy of the vertical coordinate of position so determined.

Astronomic Latitude, Longitude and Azimuth

Astronomic latitude and longitude were defined elsewhere. The *astronomic azimuth*, A, is the angle in the observers local horizon plane between the astronomic meridian of the observer and the vertical plane passing through the target point. In astronomy, azimuth is reckoned from the south point and positive

westward to the north; however, in geodesy, it is positive from the north and in the clockwise direction.

Astronomic latitudes, longitudes and azimuths are determined by observing fixed stars. Measured with a theodolite and under field conditions, astronomical latitudes and longitudes have precisions that are of the order of 0.3 arcsec. This represents about 9 m in position on the earth's surface. Field observations of astronomic azimuth have a precision of about 0.5 arcsec and this represents a linear precision of about 3 cm for a line 10 km long.

Horizontal Angles, Distances and Zenith Angles

The *horizontal angle* is the angle measured in the horizon plane of the observer between two vertical planes. It is formed by the difference in directions to the two target points which determine the vertical planes. The horizon plane is the tangent plane to the equipotential surface at the point of observation. A geodetic theodolite is used for measuring these directions. First-order direction methods typically have precisions of about 0.1 arcsec, representing a linear displacement of 5 mm for a line 10 km long but in unfavourable circumstances, and these occur frequently, atmospheric refraction associated with horizontal density gradients, can result in errors which are several times larger than this. Further errors arise from instrumental limitations and the centring of the theodolite over the station, and more typical accuracies are 0.3-0.5 arcsec for highest accuracy surveys.

Electromagnetic distance measuring devices are used to determine the *distance* between geodetic stations. Pulses of frequency-controlled light are transmitted to a retroreflector array and returned to the transmitter. The phase of the return pulse is compared with that of the emitted pulse and the observed phase shift is a measure of the distance in excess of an integral number of wavelengths of the light pulse. Intervisibility requirements and attenuation of the signal limit lengths of individual lines to about 30-50 km but the best results are obtained over shorter lines. The precision of these measurements for a line of length, L is usualy expressed as:

$$\sigma_L = (a^2 + b^2 L^2)^{1/2} \text{ cm} \qquad\qquad (6.2\text{-}1)$$

where the constant a is mainly the result of calibration errors and the centring precision of the instrument; typically a 1 cm. The length dependent term is mainly the result of atmospheric refraction; typically b 10^{-6} cm^{-1}.

The *zenith angle* is the angle down from the vertical to the target and is measured in the vertical plane of the observer and the target. The zenith angle is measured with the vertical circle of a theodolite. The principal observational limitation is atmospheric refraction. Vertical density gradients are substantially greater than horizontal density gradients and, therefore, the curvature of the lines is very much greater than in the horizontal direction. For nearly horizontal lines-of-sight and $L = 20$ km, the refraction effect is about 90 arcsec. These
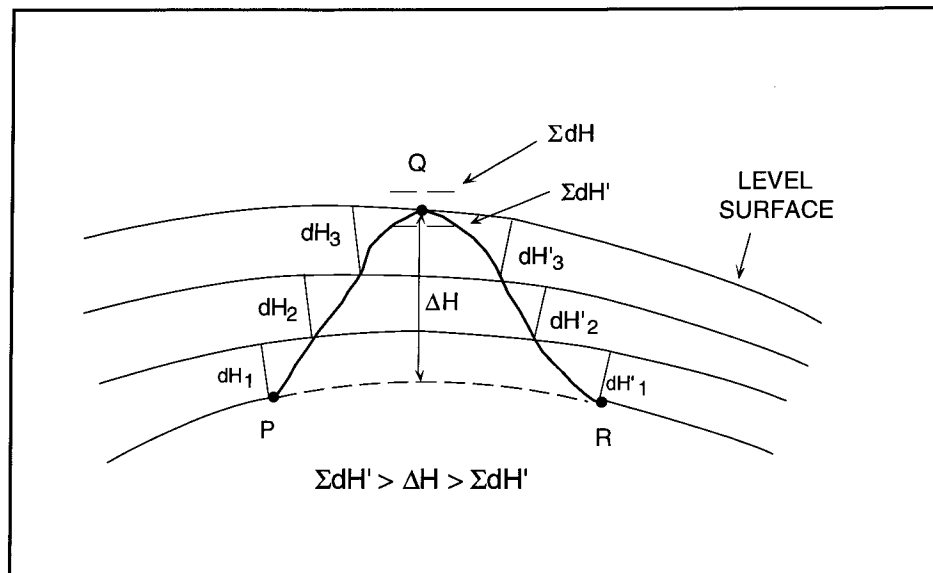
Figure 6.3-1
Height difference between widely spaced points.

effects may be reduced by observing the two reciprocal zenith angles from either end of the line simultaneously. Accuracies of about 1 arcsec can be reached for this difference when the separation between sites is less than about 10 km. Over longer distances and when the ray paths graze intervening topography, the uncertainty increases rapidly.

6.3 VERTICAL CONTROL SURVEYS

Height Differences

Heights in terrestrial geodetic surveys are determined either by geometric levelling or by trigonometrical levelling. Trigonometrically determined heights are not very accurate principally due to atmospheric refraction effects and we do not consider them further. In geometric levelling, the differences in height are determined using horizontal lines-of-sight between points in close proximity to each other. The instruments of levelling are a level and calibrated staves. The level, either a level vial or prism/mirror suspended under gravity, defines the tangent plane to the equipotential surface passing through it and the heights at which this plane intersects two staves determines the difference in height of the two stations. The distance between the staves and the level is limited to a few tens of metres for otherwise the refraction induced curvature results in the line-of-sight departing significantly from the tangent plane. Therefore, to measure height differences over longer distances, the process is repeated by successively moving the level and the staves along the line.

The accuracy of defining the horizontal surface with precision geodetic levels is about 0.1 arcsec, representing a height error of 2.5×10^{-2} mm for a 50 m separation of level and staff. For a series of measurements along a line the cumulative precision is about 0.1 mm km^{-1}, provided that the random levelling errors of the instrument represent the only source of uncertainty. This is not the case, and systematic errors, mainly caused by atmospheric refraction, become predominant over distances of 100 km or more. The accuracy of geodetic levelling over a distance L can be considered as a sum of random and systematic parts, namely:

$$\sigma_h = (a^2 L + b^2 L^2)^{1/2} \qquad (6.3\text{-}1)$$

where a and b are constants defining the magnitude of the random and systematic errors, respectively. In the highest quality levelling, where the observations are designed to reduce many of the error sources which occur and where the line has been levelled in both directions, a 1 mm $km^{-1/2}$ and b 0.2 mm km^{-1}. This represents a cumulative accuracy of 20 mm for a 100 km long line.

Height Systems

The height difference, ΔH, between two widely separated bench marks is determined by the sum of a number of height differences, ΔH_i. Consider a closed levelling loop from P to R via an elevated point Q and then back to P (Figure 6.3-1). In the absence of observational errors, the sum:

$$\sum_{P}^{Q} \Delta H_i + \sum_{Q}^{R} \Delta H_i + \sum_{R}^{P} \Delta H_i = \Delta H$$

will be zero only if the equipotential surfaces are everywhere parallel along the levelling route. This will seldom be the case for widely separated points and the elevation difference measured along this route will not, in general, vanish. Thus it follows that observed elevations are *nonunique*, that is, they depend on the levelling route taken. To avoid this nonuniqueness in definition, height differences are defined in terms of gravitational potential, dW, through the relation dW = -g dh (see 2.4-2). The difference in potential between P and Q is:

$$\Delta W_{PQ} = - \int_{P}^{Q} g\, dh$$

where g is the value of gravity along the levelling route. If point P lies on the geoid:

$$\Delta W_Q = - \int_{Geoid}^{Q} g\, dh \quad - \quad \sum_{Geoid}^{Q} g\, dh$$

Using this relationship unique heights may now be defined in a number of ways. The various height systems in use are: (1) geopotential numbers, (2) dynamic heights, (3) orthometric heights, and (4) normal heights. The *geopotential number*, C, is simply the negative potential difference between the point of interest and the geoid. Thus:

$$C_Q = \int_{Geoid}^{Q} g\,dh \quad \sum_{Geoid}^{Q} g\,dh$$

To achieve reasonable agreement with the height in metres, the unit of geopotential number is chosen as 100 m^2s^{-2}. This is called the *geopotential unit*. In this way, because g = 9.8 ms^{-2}, the geopotential numbers are about 2% smaller than the corresponding heights. If we now chose to convert geopotential numbers into units of length, we must divide them by a value of gravity, an operation which produces what are known as *dynamic heights*. The value of gravity used in this definition is usually chosen such that it represents the mean value for the region of interest.

Geopotential numbers and dynamic heights have no obvious geometric significance; neither can be plotted above the reference surface with a ruler or fixed graduation. Orthometric and normal heights overcome this problem. The *orthometric height*, H_Q, was previously defined as the linear distance along the plumbline from the geoid to the point on the surface. In practice, the orthometric height is defined as $H_Q = C_Q/g'$, where g' is the mean value of gravity along the plumbline between the geoid and Q. Actual values of gravity along the plumbline are required to evaluate g'. Since a direct measurement of gravity inside the earth is not possible, a mass distribution must be assumed with g' computed on this basis. Therefore, H cannot be determined without hypothesis. Moreover, because the level surfaces are not parallel, points of equal orthometric height are not situated on the same level surface.

Normal heights differ from the other height systems in that they use a surface called the *quasi-geoid* rather than the geoid as the reference surface. Suffice it here to say that the quasi-geoid is a surface which coincides with the geoid on the open seas, but departs from it elsewhere, typically by a few tens of centimetres and up to a few metres under high mountains. It is not an equipotential surface. Furthermore, normal heights are measured along the "plumbline" of normal gravity. They are, however, obtained without any assumptions about gravity between the reference surface and the surface point.

Satellite methods give heights relative to a reference ellipsoid. To obtain the corresponding orthometric height we need to know the geoidal undulation at the point in question. We discuss this further in Section 7.8.

Example 6.3-1

Points A, B and C are connected by precise levelling and gravity. *Calculate* the orthometric height of B and C *given* that:

Orthometric height of A.

$$H_A = 1617.492 \ m$$

Levelled elevation differences.

$$\begin{pmatrix} \Delta h_{AB} \\ \Delta h_{BC} \end{pmatrix} = \begin{pmatrix} 188.3000 \\ 253.9000 \end{pmatrix}, m$$

Mean gravity along levelling route.

$$\begin{pmatrix} \bar{g}_{AB} \\ \bar{g}_{BC} \end{pmatrix} = \begin{pmatrix} 9.80266 \\ 9.80229 \end{pmatrix}, ms^{-2}$$

Mean gravity along plumbline between geoid and B and C.

$$\begin{pmatrix} g'_{AB} \\ g'_{BC} \end{pmatrix} = \begin{pmatrix} 9.80313 \\ 9.80287 \end{pmatrix}, ms^{-2}$$

Solution

Equation (2.4-2) yields the potential differences between the points of interest:

$$\begin{pmatrix} \Delta W_{AB} \\ \Delta W_{BC} \end{pmatrix} = \begin{pmatrix} -\bar{g}_{AB} \Delta h_{AB} \\ -\bar{g}_{BC} \Delta h_{BC} \end{pmatrix} = \begin{pmatrix} -9.80266 \times 188.3 \\ -9.80229 \times 253.9 \end{pmatrix} = \begin{pmatrix} -1845.841 \\ -2488.803 \end{pmatrix}, m^2 s^{-2}$$

Dividing these by the mean value of gravity along the plumbline gives the orthometric height differences:

$$\begin{pmatrix} \Delta H_{AB} \\ \Delta H_{BC} \end{pmatrix} = \begin{pmatrix} \dfrac{-1845.841}{9.80313} \\ \dfrac{-2488.803}{9.80217} \end{pmatrix} = \begin{pmatrix} 188.29098 \\ 253.88514 \end{pmatrix}, m$$

Finally:
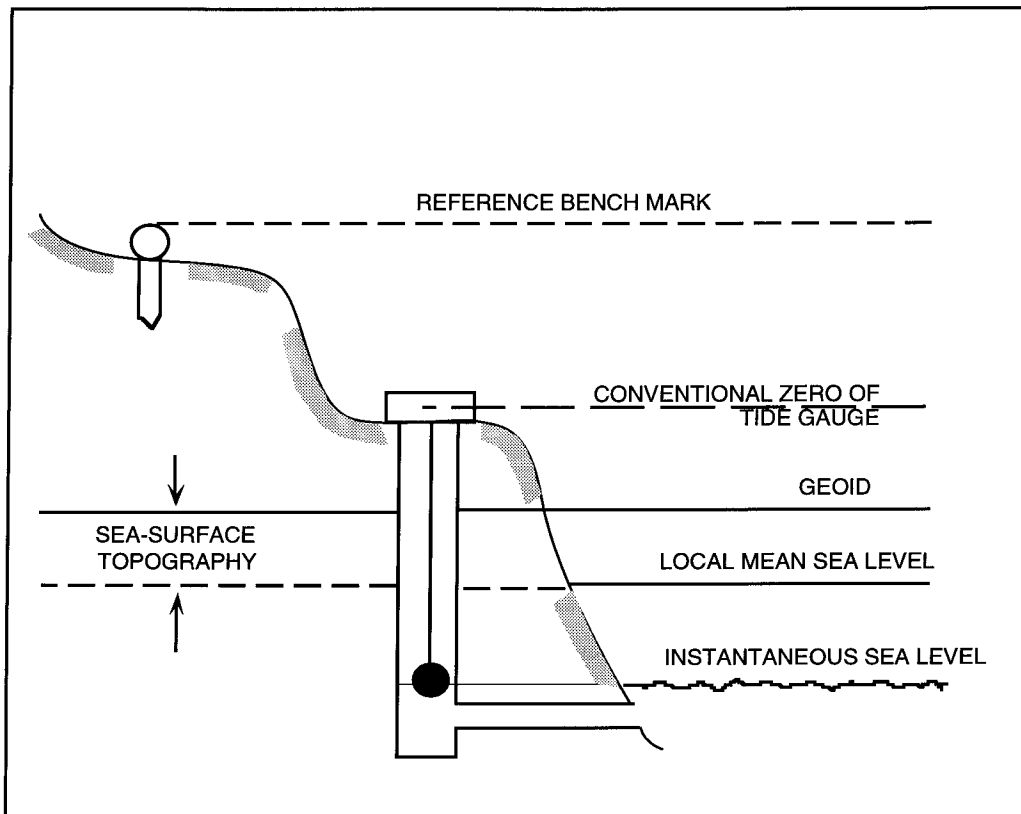
Figure 6.3-2
Local mean sealevel, instantaneous sealevel and the geoid.

$$\begin{pmatrix} H_B \\ H_C \end{pmatrix} = \begin{pmatrix} H_A \\ H_B \end{pmatrix} + \begin{pmatrix} \Delta H_{AB} \\ \Delta H_{BC} \end{pmatrix} = \begin{pmatrix} 1617.4920 \\ 1805.7830 \end{pmatrix} + \begin{pmatrix} 188.29098 \\ 253.88514 \end{pmatrix} = \begin{pmatrix} 1805.7830 \\ 2059.6681 \end{pmatrix}, m$$

Note

The heights based on levelling data alone are:

$$\begin{pmatrix} h_B \\ h_C \end{pmatrix} = \begin{pmatrix} 1805.7920 \\ 2059.6920 \end{pmatrix}, m$$

and the differences between these and the orthometric heights are:

$$\begin{pmatrix} \Delta_B \\ \Delta_C \end{pmatrix} = \begin{pmatrix} 0.0009 \\ 0.0239 \end{pmatrix}, m$$

Δ_C by far exceeds the accuracy achievable with precision levelling.
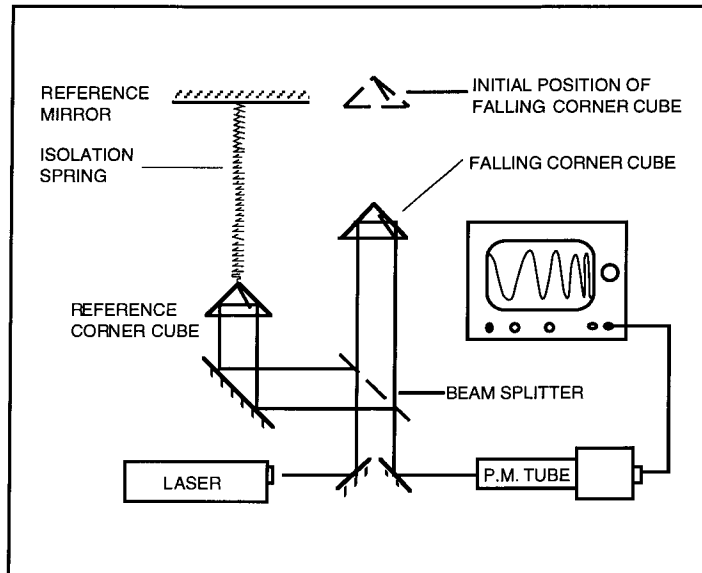
Figure 6.4-1
Principle of free-fall gravimeter.
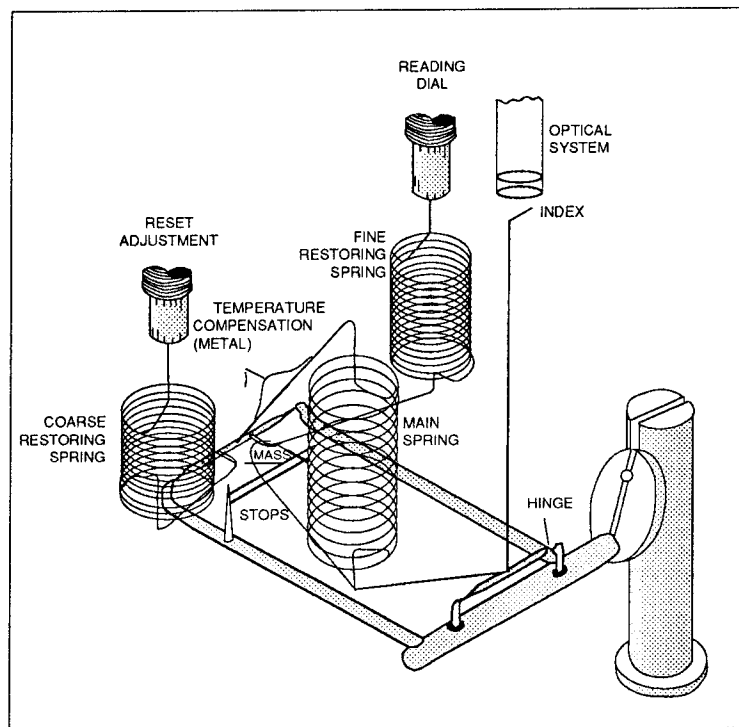
Figure 6.4-2
Principle of relative gravimeter.

Vertical Reference Datum

Mean sealevel (MSL) is the surface that has traditionally served as the zero reference datum for orthometric heights. Until several decades ago, it was believed that MSL coincided with the geoid. Hence the task of locating the position of the geoid with respect to a benchmark at the shoreline simply reduced to determining the position of local MSL. The heights of all other points of interest were then obtained from the heights of the reference benchmarks by adding the height differences along the interconnected level lines. In fact, the resultant heights above local MSL are only approximately equal to the true heights above the geoid.

MSL is established by registering and averaging the ocean's water level over longer intervals of time (1 year) using tidegauges. The variations of sealevel with time, as long as they are periodic, are eliminated by averaging the waterlevel registrations. However, since the tidegauges are often located in estuaries or harbours, they do not generally have an undisturbed link to the open oceans. Friction and bottom effects cause the readings to be falsified. Also, there are atmospheric pressure variations and currents, which produce changes in local sealevel. Averaging does not necessarily remove these effects. Thus local MSL is not the geoid. The difference between the two surfaces is called the *seasurface topography* and may amount to several tens of centimetres. Figure 6.3-2 shows the relationship the local MSL, the instantaneous sealevel and the geoid.

6.4   GRAVITY

Gravity Measurements

Gravity is measured with an instrument called a gravimeter. Two methods of gravity measurement are currently used: free-fall and force-balance. The *free-fall* method determines *absolute* gravity by measuring accurately the time versus distance of a proof mass carefully dropped in a vacuum. *Force-balance* instruments measure the force necessary to support a proof mass in the gravity field. These instruments measure *relative* gravity.

The principle of a typical modern free fall instrument is shown in Figure 6.4-1. Distance and time is measured simultaneously, with a Michelson interferometer. Accuracies of the order of 10 nms^{-2} are achievable using interferometric measurements with laser light and atomic timing devices. Transportable instruments now exist. Most relative gravimeters employ a mechanical spring in bending or torsion (Figure 6.4-2). The instruments used for most survey work are La Coste Romberg and Worden gravimeters. The accuracy of these instruments is about 1-0.1 μms^{-2}.

Gravity Networks

The values of gravity required in geodesy and geophysics must refer to a global reference system defined by values of gravity at a number of accurately surveyed gravity control points. These control points serve as a frame for subsequent detailed surveys which are performed as traverses or area surveys. We distinguish between: (1) global gravity networks; (2) regional gravity networks, and; (3) local gravity networks.

Global gravity networks, with station separations of several 100-1000 km, constitute the basic gravity reference system and are established by international cooperation. Regional gravity networks, with station separations of a few 10-100 km, generally represent the national network. Local gravity networks, with station separations of 0.1-10 km, are mostly established for geophysical exploration or geodynamical purposes.

6.5   TERRESTRIAL GEODETIC COMPUTATIONS

Below we illustrate how positions are obtained from terrestrial observations of astronomical latitude, longitude and azimuth, horizontal and vertical angles, and distance. We use the three-dimensional method because the formulas are simpler. The computations are performed without reference to an ellipsoid. It is convenient to introduce the *topocentric-horizon system* of coordinates (Figure 6.5-1). The origin of the topocentric-horizon system is the point on the earth's surface where the observer is stationed (the *topos*). The fundamental plane is the horizon and the N-axis points north. The E-axis points east and the h-axis points towards the zenith. The determinations of astronomic latitude, Φ, and longitude, Λ, establish the direction of the vertical with respect to the ITRF (Figure 6.5-1). The astronomic azimuth, A, establishes the direction of the meridian plane, while the horizontal angles, distances, s, and zenith angles, z, define the relative locations of points in the topocentric horizon system of the observer.

The coordinates of the target point in the topocentric-horizon system of the observer are (Figure 6.5-2):

$$\begin{pmatrix} \Delta N \\ \Delta E \\ \Delta h' \end{pmatrix} = s \begin{pmatrix} \cos A \ \sin z \\ \sin A \ \sin z \\ \cos z \end{pmatrix} \tag{6.5-1}$$

These components may be transformed to components in the ITRF using:
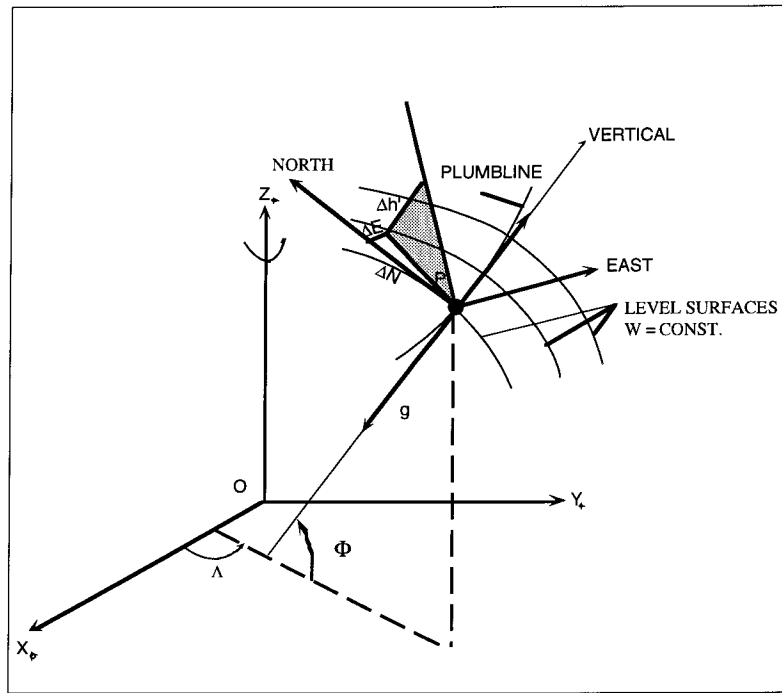
Figure 6.5-1
Topocentric-horizon system and ITRF.
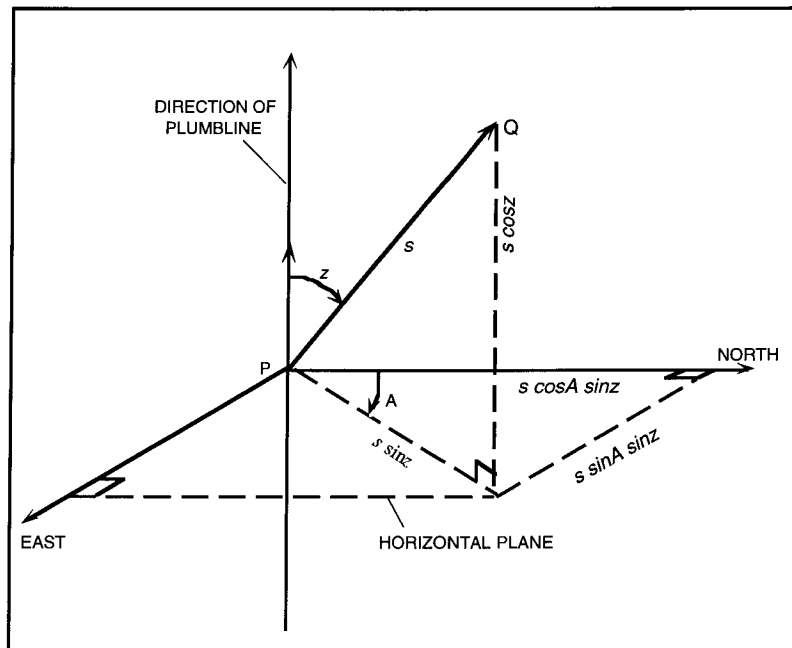
Figure 6.5-2
Relative position in the topocentric-horizon system.

$$\begin{pmatrix} \Delta X_\oplus \\ \Delta Y_\oplus \\ \Delta Z_\oplus \end{pmatrix} = \mathbf{C} \begin{pmatrix} \Delta N \\ \Delta E \\ \Delta h' \end{pmatrix} \qquad (6.5\text{-}2)$$

where,

$$\mathbf{C} = \begin{pmatrix} -\sin\Phi\cos\Lambda & -\sin\Lambda & \cos\Phi\cos\Lambda \\ -\sin\Phi\sin\Lambda & \cos\Lambda & \cos\Phi\sin\Lambda \\ \cos\Phi & 0 & \sin\Phi \end{pmatrix} \qquad (6.5\text{-}3)$$

Note

The geodetic horizon is sometimes used. This is tangential to the reference ellipsoid directly below the observer. The geodetic and topocentric horizons are related by the deflection of the vertical. To transform geodetic horizon coordinates to the ITRF we use the same equations but with geodetic latitude and longitude.

Example 6.5-1

Figure 6.5-3 shows the angular observations made with a theodolite at point B to connect the points A and C. *Compute* the ITRF coordinates of point C *given* that:

ITRF coordinates of B.

$$\mathbf{X_B} = \begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix} = \begin{pmatrix} -3\ 894\ 420.176 \\ 3\ 847\ 223.501 \\ -3\ 262\ 482.840 \end{pmatrix}, \text{ metres}$$

Astronomic latitude and longitude of B.

$$\begin{pmatrix} \Phi \\ \Lambda \end{pmatrix} = \begin{pmatrix} -30°\ 57'\ 46\rlap{.}''08 \\ 135°\ 20'\ 53\rlap{.}''19 \end{pmatrix}$$

Astronomic azimuth BA.

$$A_{BA} = 270°\ 29'\ 11\rlap{.}''99$$

Zenith angle at B to C.

$$z = 90°\ 06'\ 14\rlap{.}''9$$
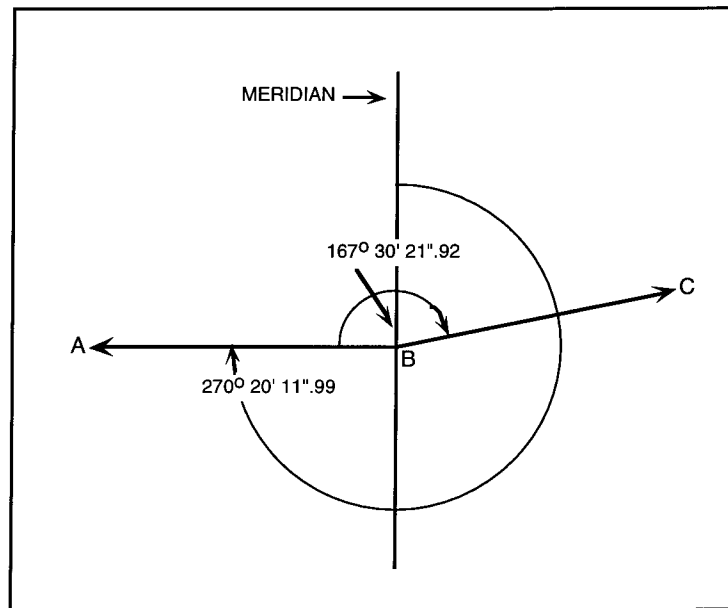
Figure 6.5-3
Horizontal angles and azimuth.

Slant distance BC.

$$s = 23\ 267.23\ \text{m}$$

Solution

We first compute the azimuth of the line BC from A_{BA} and the measured angle ABC:

$$A_{BC} = A_{BA} + 167°\ 39'\ 21\overset{''}{.}92 = 78°\ 08'\ 33\overset{''}{.}91$$

Now, we compute the components of the line BC in the ITRF from (6.5-2):

$$\Delta\mathbf{X}_{BC} = \begin{pmatrix} \Delta X_\oplus \\ \Delta Y_\oplus \\ \Delta Z_\oplus \end{pmatrix} = s \begin{pmatrix} -\sin z \sin A \sin\Lambda - \sin z \cos A \sin\Phi \cos\Lambda + \cos z \cos\Phi \cos\Lambda \\ -\sin z \sin A \cos\Lambda - \sin z \cos A \sin\Phi \sin\Lambda + \cos z \cos\Phi \sin\Lambda \\ \sin z \cos A \cos\Phi + \cos z \sin\Phi \end{pmatrix}$$

$$= \begin{pmatrix} -17\ 727.176 \\ -14\ 495.726 \\ 4\ 121.303 \end{pmatrix},\ \text{m}$$

Finally, we compute the ITRF coordinates of C, ie. $\mathbf{X}_C = \mathbf{X}_B + \Delta\mathbf{X}_{BC}$:

$$
\begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix} = \begin{pmatrix} \text{-3 894 420.176} \\ \text{3 847 223.501} \\ \text{-3 262 482.840} \end{pmatrix} + \begin{pmatrix} \text{-17 727.176} \\ \text{-14 495.726} \\ \text{4 121.303} \end{pmatrix} = \begin{pmatrix} \text{-3 912 147.352} \\ \text{3 832 727.775} \\ \text{-3 258 361.537} \end{pmatrix}, m
$$

Example 6.5-2

Convert the ITRF coordinates of point C obtained in Example 6.5-1 to ϕ, λ and h. *Take* a = 6 378 160 m and e^2 = 0.006 694 541 855.

Solution

We have:

$$
\begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix} = \begin{pmatrix} (\nu+h)\cos\phi\,\cos\lambda \\ (\nu+h)\cos\phi\,\sin\lambda \\ [\nu(1-e^2)+h]\sin\phi \end{pmatrix}
\tag{5.2-8}
$$

Dropping subscripts:

$$
\lambda = \tan^{-1}\frac{Y}{X} = 135°\,35'\,15\rlap{.}''0619
$$

To obtain ϕ and h, we write:

$$
p = \sqrt{X^2+Y^2} = (\nu+h)\cos\phi
$$

Hence:

$$
h = \frac{p}{\cos\phi} - \nu
\tag{6.5-4}
$$

Manipulation of (5.2-8) and dividing by p gives:

$$
\frac{Z}{p} = (1-e^2\frac{\nu}{\nu+h})\,\tan\phi
$$

Accordingly:

$$
\phi = \tan^{-1}[\frac{Z}{p}\,(1-e^2\frac{\nu}{\nu+h})^{-1}]
\tag{6.5-5}
$$

Equations (6.5-4) and (6.5-5) are now solved iteratively for ϕ and h. As a first approximation, we set h = 0 in (6.5-5), obtaining:

$$\phi_{(1)} = \tan^{-1}[\frac{Z}{p}(1-e^2)^{-1}] = -30°919711$$

Using $\phi_{(1)}$, we now compute an approximate value $v_{(1)}$ using (5.2-6):

$$v_{(1)} = \frac{a}{\sqrt{1-e^2\sin^2\phi}} = 6\ 383\ 804.327\ m$$

Equation (6.5-4) gives $h_{(1)} = 176.418$ m.

As a second approximation, we set $h = h_{(1)}$ in (6.5-5), obtaining:

$$\phi_{(2)} = \tan^{-1}[\frac{Z}{p}(1-e^2\frac{v_{(1)}}{v_{(1)}+h_{(1)}})^{-1}] = -30°91970641$$

Using $\phi_{(2)}$, improved values of v and h are found and the process is repeated until ϕ and h remain practically constant.

We get $v_{(2)} = 6\ 383\ 804.327$ m and $h_{(2)} = 176.101$ m.

The third approximation gives:

$$\phi_{(3)} = -30°91970643, \quad v_{(3)} = 6\ 383\ 804.327\ m, \quad h_{(3)} = 176.106\ m.$$

And, the fourth approximation:

$$\phi_{(4)} = -30°91970643, \quad v_{(4)} = 6\ 383\ 804.327\ m, \quad h_{(4)} = 176.106\ m.$$

which has converged.

Hence, the required geodetic coordinates are:

$$\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} -30° 55' 10.''9431 \\ 135° 35' 15.''0619 \\ 176.106\ m \end{pmatrix}$$

Example 6.5-3

Compute the forward and reverse azimuth of the line AB given that:

Geodetic coordinates of A.

$$\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} -35^\circ \ 38' \ 10.''5140 \\ 148^\circ \ 56' \ 21.''5252 \\ 1349.893 \ m \end{pmatrix}$$

Geodetic coordinates of B.

$$\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} -35^\circ \ 23' \ 54.''4632 \\ 148^\circ \ 58' \ 39.''6605 \\ 674.375 \ m \end{pmatrix}$$

Use $a = 6 \ 378 \ 137.0 \ m$ and $\frac{1}{f} = 298.257$

<u>Solution</u>

Equations (5.2-8) give the ITRF coordinates of A and B ie.:

$$\begin{pmatrix} X_A \\ Y_A \\ Z_A \end{pmatrix} = \begin{pmatrix} -4 \ 446 \ 476.915 \\ 2 \ 678 \ 127.000 \\ -3 \ 696 \ 251.423 \end{pmatrix}, m$$

$$\begin{pmatrix} X_B \\ Y_B \\ Z_B \end{pmatrix} = \begin{pmatrix} -4 \ 460 \ 935.259 \\ 2 \ 682 \ 765.697 \\ -3 \ 674 \ 381.388 \end{pmatrix}, m$$

The components of the vectors AB and BA are then:

$$\begin{pmatrix} \Delta X_{AB} \\ \Delta Y_{AB} \\ \Delta Z_{AB} \end{pmatrix} = \begin{pmatrix} -14 \ 458.344 \\ 4 \ 638.697 \\ 21 \ 870.035 \end{pmatrix}, m$$

$$\begin{pmatrix} \Delta X_{BA} \\ \Delta Y_{BA} \\ \Delta Z_{BA} \end{pmatrix} = \begin{pmatrix} 14 \ 458.344 \\ -4 \ 638.697 \\ -21 \ 870.035 \end{pmatrix}, m$$

Since:

$$\begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} = C \begin{pmatrix} \Delta N \\ \Delta E \\ \Delta h' \end{pmatrix}$$

We have:

$$\begin{pmatrix} \Delta N \\ \Delta E \\ \Delta h \end{pmatrix} = \mathbf{C}^{-1} \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z' \end{pmatrix} = \begin{pmatrix} -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi \end{pmatrix} \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z' \end{pmatrix}$$

Using 10-digit arithmetic to evaluate $\mathbf{C}^{-1}$ for both A and B we get:

$$\begin{pmatrix} \Delta N_{AB} \\ \Delta E_{AB} \\ \Delta h_{AB} \end{pmatrix} = \begin{pmatrix} 26\ 385.071 \\ 3\ 486.114 \\ -731.221 \end{pmatrix}, m$$

$$\begin{pmatrix} \Delta N_{BA} \\ \Delta E_{BA} \\ \Delta h_{BA} \end{pmatrix} = \begin{pmatrix} -26\ 389.229 \\ -3\ 476.216 \\ 619.810 \end{pmatrix}, m$$

As a check on the calculations we compute the distances AB and BA. We find that AB = BA = 26 624.418 m.

The forward and reverse azimuths of AB are:

$$A_{AB} = \tan^{-1}\frac{\Delta E_{AB}}{\Delta N_{AB}} = \tan^{-1}\frac{3\ 486.114}{26\ 385.071} = \quad 7°\ 31'\ 35''.69$$

$$A_{BA} = \tan^{-1}\frac{\Delta E_{BA}}{\Delta N_{BA}} = \tan^{-1}\frac{-3\ 476.216}{-26\ 389.229} = 187°\ 30'\ 15''.43$$

Observe that $A_{BA} \neq 180° + A_{AB}$. This is because the meridians through A and B are not parallel; rather, they converge towards the South Pole, in this instance. The difference 1' 20''.26 is known as the *meridian convergence*. This increases with increasing latitude and separation in longitude.

REFERENCES

BOMFORD, G., 1985. *Geodesy* (4th edition), Clarendon Press, Oxford, 855pp.

HEISKANEN, W.A. & MORITZ, H., 1967. *Physical Geodesy*, W.H. Freeman, San Francisco, 364pp.

LAMBECK, K., 1988. *Geophysical Geodesy*, Clarendon Press, Oxford, 718pp.

TORGE, W., 1991. *Geodesy* (2nd edition), Walter de Gruyter, Berlin, 264pp.

EXERCISES

6.1   Briefly describe the geodetic surveying methods of triangulation, trilateration and traversing. Give the strengths and weaknesses of each method.

6.2   Distinguish between the projection and three-dimensional methods of geodetic computation. Which is more accurate in practise. Give reasons for your answer.

6.3   Name two terrestrial geodetic methods for obtaining height. Which is more accurate? Give reasons for your answer?

6.4   What is a datum point?

6.5   What is an azimuth? How is it determined and what accuracy is achievable?

6.6   Distinguish between a geodetic and astronomic azimuth.

6.7   Show that an azimuth error of 0.3 arcsec represents a linear precision of about 3 cm for a line 10 km long.

6.8   How might satellite methods be used to establish an azimuth?

6.9   Why are horizontal angles more accurate than vertical angles?

6.10   What are reciprocal vertical angles? Why are they used?

6.11   Geodetic levelling is one of the most accurate geodetic measurements which can be made. Discuss this statement.

6.12   Distinguish between dynamic, orthometric and normal heights.

6.13   Using a diagram to illustrate your answer show that levelling alone does not give orthometric height.

6.14   Mean sealevel need not coincide with the geoid. Discuss this statement.

6.15   Compute the latitude, longitude and ellipsoidal height of point B given that:

ITRF coordinates of point A.

$$\begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix} = \begin{pmatrix} \text{-444 6476.915} \\ \text{+267 8127.000} \\ \text{-369 6251.423} \end{pmatrix}, m$$

Azimuth, zenith angle, slope distance of AB.

$$\begin{pmatrix} \alpha \\ z \\ s \end{pmatrix} = \begin{pmatrix} 48° \ 10' \ 00.''32 \\ 102° \ 01' \ 26.''6 \\ 1926.286 \ m \end{pmatrix}$$

Take a = 6 378 137.0 m and $\frac{1}{f}$ = 298.257.

Answer:

$$\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} \text{-35}° \ 37' \ 29.''7450 \\ 148° \ 57' \ 17.''3027 \\ 948.883 \ m \end{pmatrix}$$

6.16   The geocentric Cartesian coordinates of an observer on the earth are:

$$\begin{pmatrix} X_\oplus \\ Y_\oplus \\ Z_\oplus \end{pmatrix} = \begin{pmatrix} \text{-4460935.259} \\ 2682765.697 \\ \text{-3674381.388} \end{pmatrix}, m$$

Calculate the corresponding geodetic coordinates ϕ, λ and h. Take a = 6 378 137.0 m and $\frac{1}{f}$ = 298.257.

Answer:

$$
\begin{pmatrix} \phi \\ \lambda \\ h \end{pmatrix} = \begin{pmatrix} -35^\circ\ 23'\ 54.''4632 \\ 148^\circ\ 58'\ 39.''6605 \\ 674.375\ m \end{pmatrix}
$$

6.17 What is meridian convergence? What happens to it at (a) the equator and (b) near the poles?

7 Space Geodetic Methods

7.1 INTRODUCTION

The space geodetic methods currently used for precise positioning and earth gravity field determination are: (1) laser ranging to artificial satellites, (2) biased range measurements to the satellites of the Global Positioning System, (3) radio interferometric observations of extragalactic sources, and (4) altimeter measurements from a satellite to the ocean surface. The altimetric satellites orbit the earth at about 1 000 km altitude, the laser ranging satellites at about 6 000 km and the GPS satellites at about 20 000 km, while the extragalactic radio sources are located at infinity. We briefly describe these systems here. Also, we show how, in principle, they are used for determining geodetic parameters.

7.2 SATELLITE LASER RANGING

Satellite laser range (SLR) observations consist of the measurement of the time taken for a short energy pulse to travel from the transmitter at the ground station to the satellite and back. The time delay between transmission and reception is thus a measure of twice the distance to the satellite. The pulse is generated by a laser, focussed onto the satellite, and reflected back to the station by an array of retroreflectors mounted on the spacecraft. Generally, the reflectors are uniformly
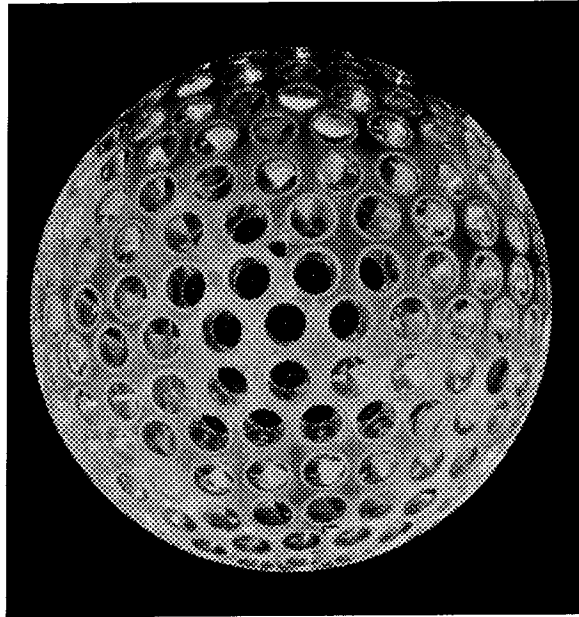
Figure 7.2-1
*Laser geodynamics satellite (LAGEOS): semi-major axis, 12 270 km;
eccentricity, 0.004; inclination, 109.9°; cross-sectional area, 0.283 m²; mass,
411 kg.*

distributed over a spherical outer shell. This is the case for the two LAGEOS satellites which were launched solely as a target for lasers (Figure 7.2-1). These satellites are quite small (60 cm diameter) but heavy (411 kg).

The transmitted pulse length of current systems is approximately 200 picoseconds, which is equivalent to 6 cm of two-way travel time. Single-shot system precision is well below 1 cm for the technology used in the most precise laser systems which have been developed todate. Systematic error sources, such as electronic delays within the transmitter and receiver, and epoch timing errors currently vary from about 2-10 mm. Atmospheric path delay uncertainties amount to about 5 mm. Two-colour laser ranging systems now under development should reduce the atmospheric refraction uncertainties.

The three-dimensional positioning accuracy currently achievable with SLR is about 1 cm. Several weeks or months of data are required to obtain this accuracy. This is because 40-60 satellite passes are needed in cloudless weather conditions. Currently, there are more than 20 laser stations operating at various locations around the globe. Many of these stations are clustered in Europe and North America; two are located in Australia, one near Geralton (WA) and the other at Orroral (ACT). Both fixed and mobile systems have been developed.
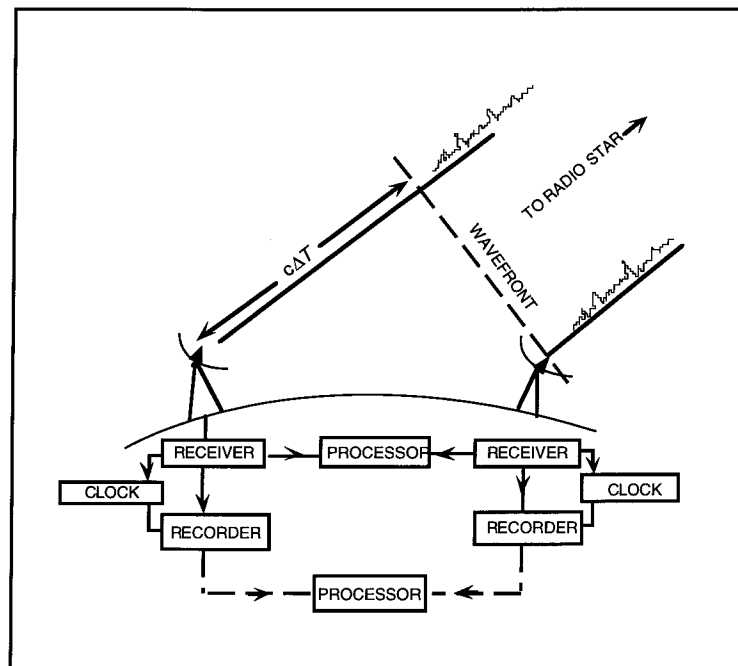
Figure 7.3-1

Schematic illustration of VLBI observations. The signals are recorded independently at the two antennas, together with clock signals, and later brought together in the processor.

7.3   VERY LONG BASELINE INTERFEROMETRY

In Very Long Baseline Interferometry (VLBI), an extragalactic radio source is observed at frequencies centred near 8.3 GHz and 2.3 GHz simultaneously from two or more radio telescopes with antennas of 3-100 m in diameter (Figure 7.3-1). The radio source is typically a quasar. The received signals will be in phase instantaneously at any two telescopes only when their distances to the remote source are identical. In general, the difference in arrival time, $\Delta\tau$, is proportional to the baseline length, L, joining the two telescopes and to the cosine of the angle between the baseline and the source, α:

$$L = c\Delta\tau \cos\alpha \qquad\qquad (7.3\text{-}1)$$

In a typical 24-hour observing session, as many as 200 observations are made to more than a score of radio sources, permitting estimates to be made of the components of the baseline vector L to within a centimetre. Hydrogen maser clocks and high-data-rate recorders are essential to record the radio frequency signal from selected radio sources at each radio telescope. Simultaneous data from several telescopes are processed by dedicated digital autocorrelators to determine instantaneous signal delays and their rate-of-change with time. As

with SLR systems, existing VLBI telescopes are concentrated in the northern hemisphere. Australia hosts three telescopes with geodetic VLBI capability: Tidbinbilla (ACT), Hobart, and the Australia Telescope near Narrabri (NSW).

7.4   THE GLOBAL POSITIONING SYSTEM

SLR and VLBI have major limitations when rapid, moderately accurate, positions are required. The laser range measurements require clear weather and several weeks of data to produce a position determination which is accurate enough. VLBI systems can operate in all weather and can provide accurate results from just 24-hours of observations. However, mobile VLBI is expensive and not particularly portable. These limitations have been largely overcome with GPS. A total of 21 satellites plus three spares are planned to be in operation (Figure 7.4-1). They will orbit at an altitude of about 20 000 km in six orbital planes with 12-hour periods, enabling simultaneous observation of four or more satellites in virtualy all parts of the globe.

The system, which is now almost fully configured, has been used by geodesists for some years. The GPS satellites are observed continuously from a global network of permanent tracking stations and the orbits are computed at regular and frequent intervals. The satellite positions and velocities are extrapolated ahead of time and this information is transmitted from the satellite to the observer at the time of observation. These orbits are accurate to about 10-40 m. More accurate orbits may be obtained from additional GPS data taken simultaneously from sites whose positions are well known from independent measurements such as VLBI or SLR.

As with SLR, distance information in GPS is based on the travel time, τ, of a satellite signal, obtained by measuring the difference between the transmit time, t_T, and receive time, t_R, at the GPS receiver of a special ranging code. If we ignore transmission media effects on the velocity-of-light, c, and timing errors, then the true range, ρ, between the satellite and receiver is:

$$\rho = c\tau = c(t_R\text{-}t_T) \tag{7.4-1}$$

This assumes that time kept at both the satellite and the receivers are synchronized with a very high degree of precision. The satellites are equipped with caesium and rubidium clocks and these are controlled by comparing them with a ground station master clock. Clocks at the ground receivers may not be synchronized with the satellite clocks so that each range measurement may be in error by a constant amount. Also, the receiver clocks drift with time and the offset between receiver and transmitter clocks does not remain constant. The actual range measurement is therefore referred to as the *pseudorange*, R:

$$R = \rho + c(\Delta t_R\text{-}\Delta t_T) \tag{7.4-2}$$
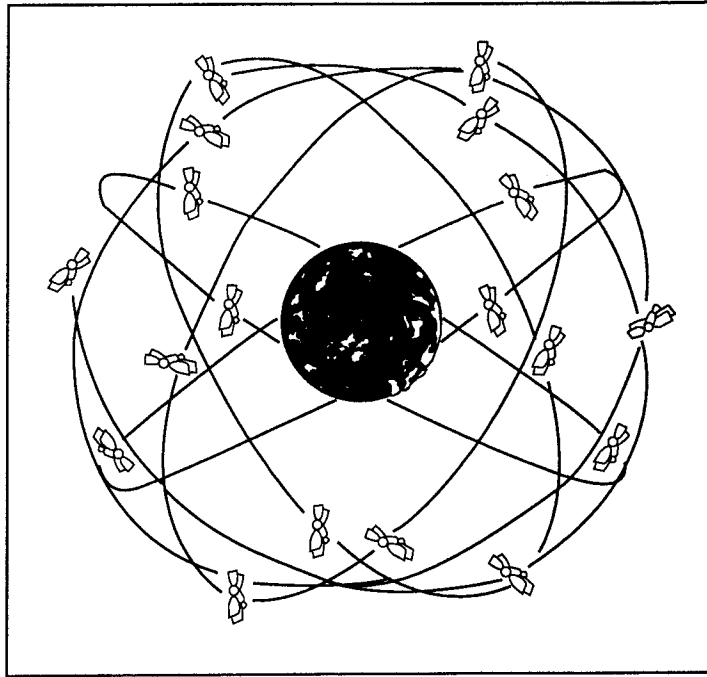
Figure 7.4-1

GPS constellation: 18 satellites, 20 000 km altitude, 12 h orbits.
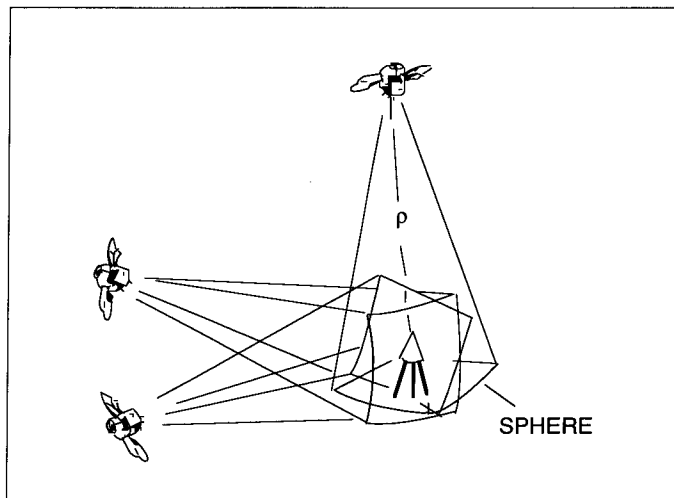
Figure 7.4-2

Principle of satellite positioning. The ranges to three satellites are measured simultaneously. The locations of the satellites in an earth-fixed reference coordinate system are known. The observer's position is determined by the intersection of three spheres, of radius equal to the measured range and centred on the satellite.

where Δt_R is the receiver clock offset from GPS time and Δt_T is the corresponding satellite clock offset.

The GPS measurements of pseudorange can be made by means of special codes transmitted by the satellites. Each satellite transmits two carrier waves of about 1.2 and 1.6 GHz that are modulated with range codes that define a sequence of accurate time marks that are also produced within the receiver. Hence the phase shifts can be measured and this determines the pseudorange. Two codes are available; a precision code with 30 m wavelength and a coarse code with 300 m wavelength. The range precisions of these two codes are about 1 and 10 m respectively.

Geometrically, the observed distance constrains the receiver to lie on a sphere centred on the satellite and, in the absence of a receiver clock offset error, three such observations to well-spaced satellite positions locate the position of the receiver at the intersection of the three spheres (Figure 7.4-2). If a fourth satellite is observed at the same time then, assuming that the satellite clock keeps GPST, the station clock offset can be determined as well. The actual procedures used for positioning with GPS and SLR differ only in detail from this outline.

Example 7.4-1

Figure 7.4-3 illustrates the geometry of earth-based satellite observations. *Show* that small changes (dX,dY,dZ) in station coordinates (X,Y,Z) and small changes (dx,dy,dz) in satellite position (x,y,z) produce a change in the range to the satellite, dρ, which is given by:

$$d\rho = \frac{x\text{-}X}{\rho}(dx\text{-}dX) + \frac{y\text{-}Y}{\rho}(dy\text{-}dY) + \frac{z\text{-}Z}{\rho}(dz\text{-}dZ)$$

Solution

We first express the range to the satellite in terms of station and satellite position:

$$\rho = \sqrt{(x\text{-}X)^2 + (y\text{-}Y)^2 + (z\text{-}Z)^2}$$

This is a function of four variables, that is:

$$\rho = f(x,y,z,X,Y,Z)$$

Accordingly:

$$d\rho = \frac{\partial \rho}{\partial x}dx + \frac{\partial \rho}{\partial y}dy + \frac{\partial \rho}{\partial z}dz + \frac{\partial \rho}{\partial X}dX + \frac{\partial \rho}{\partial Y}dY + \frac{\partial \rho}{\partial Z}dZ$$
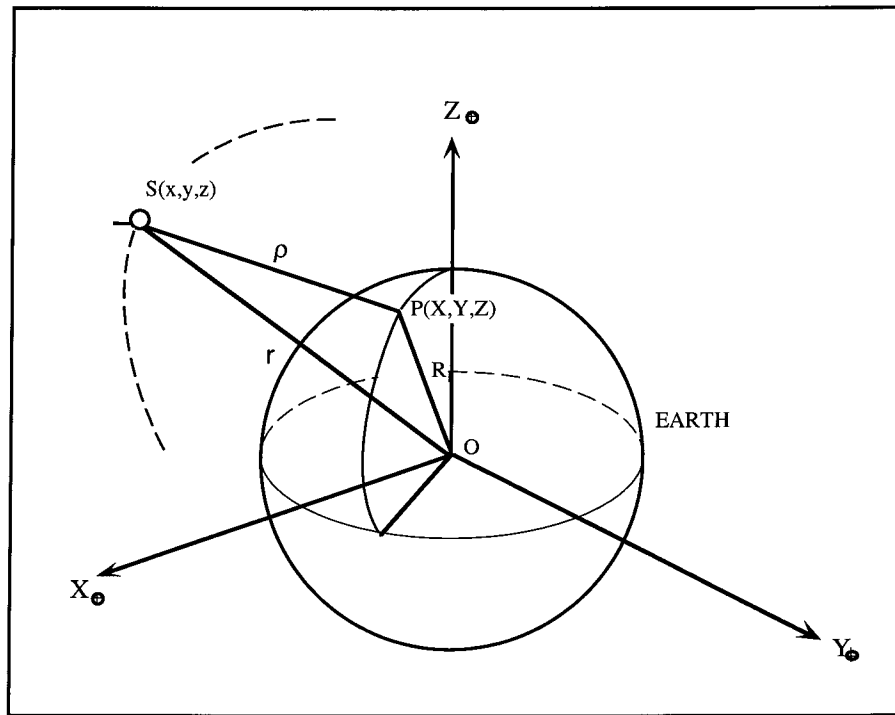
Figure 7.4-3
Geometry of earth-based satellite observations.

We square ρ before differentiating to obtain $d\rho$:

$$\rho^2 = (x\text{-}X)^2 + (y\text{-}Y)^2 + (z\text{-}Z)^2$$

Then:

$$2\rho d\rho = 2(x\text{-}X)(dx\text{-}dX) + 2(y\text{-}Y)(dy\text{-}dY) + 2(z\text{-}Z)(dz\text{-}dZ)$$

Finally:

$$d\rho = \frac{x\text{-}X}{\rho}(dx\text{-}dX) + \frac{y\text{-}Y}{\rho}(dy\text{-}dY) + \frac{z\text{-}Z}{\rho}(dz\text{-}dZ) \qquad (7.4\text{-}3)$$

In geodesy we call (7.4-3) an *observation equation*. It is used to improve parameters of interest which are only approximately known.

Example 7.4-2

An observer locates his position in an earth-fixed reference frame by making simultaneous range measurements to three satellites. The range measurements are:
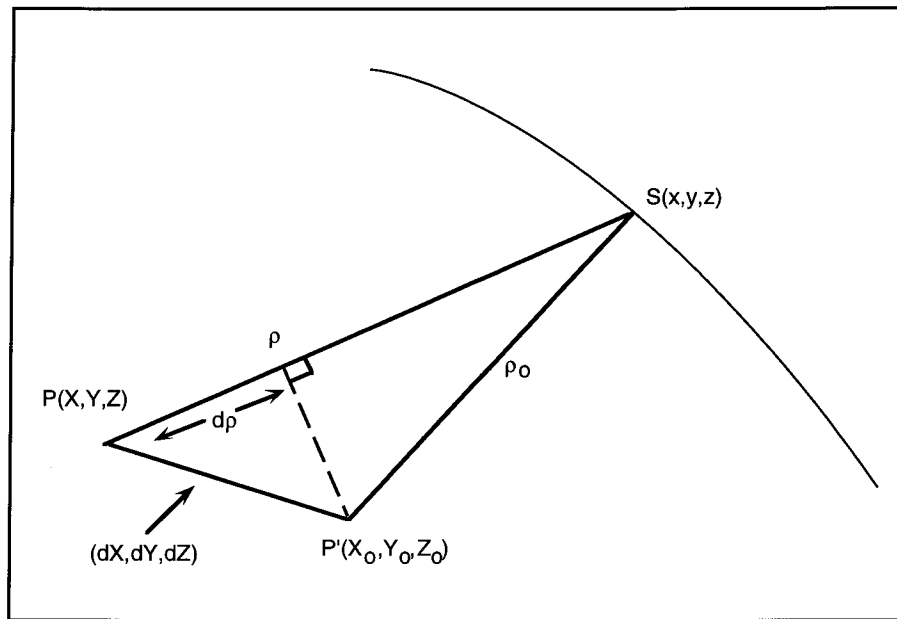
Figure 7.4-4
Determining an observers exact location from satellite range measurements.
$P'(X_o,Y_o,Z_o)$ is the observer's approximate position, $P(X,Y,Z)$ is his exact
position, ρ is measured range and ρ_o is the distance between the satellite and
P'.

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = \begin{pmatrix} 20\ 913\ 888.3 \\ 21\ 792\ 394.1 \\ 21\ 365\ 309.1 \end{pmatrix}, m$$

The satellite positions in the earth-fixed system at the instant of observation are:

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} = \begin{pmatrix} -24\ 490\ 981.3 & 3\ 022\ 488.5 & -10\ 127\ 076.0 \\ -12\ 609\ 197.8 & -2\ 504\ 553.9 & -23\ 183\ 608.5 \\ -9\ 561\ 367.8 & 10\ 473\ 490.4 & -22\ 758\ 606.0 \end{pmatrix}, m$$

The observers approximate coordinates are:

$$\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} = \begin{pmatrix} -4\ 648\ 500 \\ 2\ 546\ 490 \\ -3\ 536\ 140 \end{pmatrix}, m$$

Assuming that the range measurements and satellite coordinates are perfect *calculate* the observers exact location.

Solution

Since dx = dy = dz = 0 in (7.4-3), we may write for each range measurement:
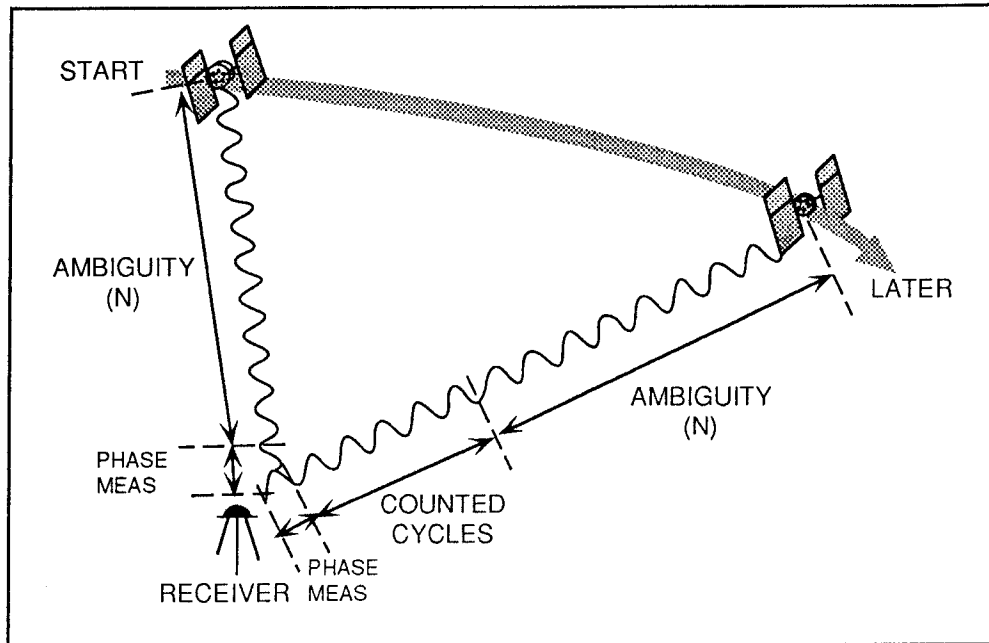
Figure 7.4-5
GPS phase measurements.

$$d\rho = -\frac{x-X_0}{\rho} dX - \frac{y-Y_0}{\rho} dY - \frac{z-Z_0}{\rho} dZ = \rho - \rho_0$$

where $\rho_0 = \sqrt{(x-X_0)^2 + (y-Y_0)^2 + (z-Z_0)^2}$ is the distance to the satellite computed from the approximate position of the observer and the exact coordinates of the satellite. Figure 7.4-4 depicts the geometry of the problem.

Now:

$$\rho_{01} = \sqrt{(x_1-X_0)^2 + (y_1-Y_0)^2 + (z_1-Z_0)^2} = 20\ 913\ 896.72\ \text{m}$$

Similarly:

$$\rho_{02} = 21\ 792\ 401.71\ \text{m} \quad \text{and} \quad \rho_{03} = 21\ 365\ 317.80\ \text{m}$$

The observation equation for the first satellite is:

$$-\frac{x_1-X_0}{\rho_1} dX - \frac{y_1-Y_0}{\rho_1} dY - \frac{z_1-Z_0}{\rho_1} dZ = \rho_1 - \rho_{01}$$

Substitution gives:

$$0.9488\ dX - 0.0228\ dY + 0.3151\ dZ = -8.42\ \text{m}$$

For ρ_2 and ρ_3 we get:

$$0.3653 \, dX + 0.2318 \, dY + 0.9016 \, dZ = -7.61 \text{ m}$$

$$0.2299 \, dX - 0.3710 \, dY + 0.8997 \, dZ = -8.70 \text{ m}$$

Expressing this in matrix form gives:

$$\begin{pmatrix} 0.9488 & -0.0228 & 0.3151 \\ 0.3653 & 0.2318 & 0.9016 \\ 0.2299 & -0.3710 & 0.8997 \end{pmatrix} \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix} = \begin{pmatrix} -8.42 \\ -7.61 \\ -8.70 \end{pmatrix}$$

Solving for dX, dY and dZ:

$$\begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix} = \begin{pmatrix} -6.60 \\ 3.31 \\ -6.62 \end{pmatrix}, \text{ m}$$

Finally:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix} = \begin{pmatrix} -4\,648\,506.60 \\ 2\,546\,493.31 \\ -3\,536\,146.62 \end{pmatrix}, \text{ m}$$

Note

We need to observe a fourth satellite if we admit a receiver clock error, dt. However, the above equations must first be modified to include dt. Assuming that $\Delta t_T = 0$ in (7.4-2) and using (7.4-3) we may write:

$$-\frac{x-X_0}{\rho} \, dX - \frac{y-Y_0}{\rho} \, dY - \frac{z-Z_0}{\rho} \, dZ - cdt = \rho - \rho_0$$

Example 7.4-3

The simultaneous range to a fourth satellite in the above example is $\rho_4 = 19\,878$ 167.2 m. *Calculate* the observer's position and the error of his clock if the satellite coordinates in the earth-fixed reference frame at the instant of observation are:

$$\begin{pmatrix} x_4 \\ y_4 \\ z_4 \end{pmatrix} = \begin{pmatrix} -17\,965\,588.1 \\ 12\,319\,810.7 \\ -14\,594\,136.4 \end{pmatrix}, \text{ m}$$

Solution

The appropriate observation equations are:

$$\begin{pmatrix} 0.9488 & -0.0228 & 0.3151 & 1 \\ 0.3653 & 0.2318 & 0.9016 & 1 \\ 0.2299 & -0.3710 & 0.8997 & 1 \\ 0.6699 & -0.4917 & 0.5563 & 1 \end{pmatrix} \begin{pmatrix} dX \\ dY \\ dZ \\ cdt \end{pmatrix} = \begin{pmatrix} -8.42 \\ -7.61 \\ -8.70 \\ -9.71 \end{pmatrix}$$

Solving for dX, dY, dZ and dt gives:

$$\begin{pmatrix} dX \\ dY \\ dZ \\ cdt \end{pmatrix} = \begin{pmatrix} -6.39 \\ 3.26 \\ -6.39 \\ 0.27 \end{pmatrix}, m$$

The station positions are:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix} = \begin{pmatrix} -4\ 648\ 506.39 \\ 2\ 546\ 493.26 \\ -3\ 536\ 146.39 \end{pmatrix}, m$$

We take c = 299 792 458 ms^{-1}. This gives:

$$dt = \frac{0.27}{c} \approx 1 \text{ ns}$$

It is also possible to obtain distance information from phase measurements on GPS the carrier signal, keeping track of the number of cycles after signal acquisition. Assuming perfect clocks and ignoring propagation effects:

$$\rho = N\lambda + \phi\lambda = \frac{c}{f}(N+\phi) \tag{7.4-4}$$

where N is the integer number of carrier wavelengths at signal acquisition, which is initially unknown, ϕ is the phase in cycles, λ is the wavelength and f is the frequency (Figure 7.4-5). The phase measurement can give a precision for the distance of about 1 cm or better. Also, since the wavelength of the carrier is considerably shorter than the wavelength of the code, the resulting distance measurement, though ambiguous by the initial number of wavelengths, is considerably more precise than the pseudorange measurement.

The GPS satellites may be used in the relative positioning mode to determine 50 km long baseline vectors with an accuracy of 1 cm horizontally and 1-3 cm vertically. For short baselines (<5 km) where atmospheric conditions at each site may be considered uniform, relative elevations and baseline lengths may be determined to subcentimetre precision, the main limitation here being
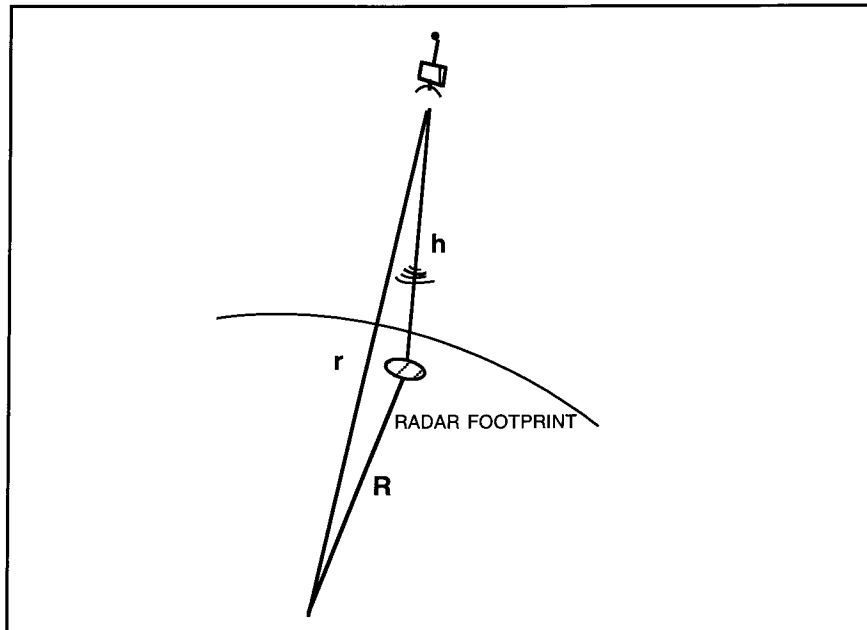
Figure 7.5-1
*Geometry for altimeter observations. The observed quantity is the satellite
height, h, above a point on the ocean surface directly below the spacecraft.*

instrumental system noise. The accuracy of longer lines is limited ultimately by
the accuracy to which the orbits of the GPS satellites are known and by the
effectiveness of processing procedures to eliminate imperfections in the data.
The GPS baseline uncertainty may be expressed in terms of baseline length, L
in km, as:

$$\sigma_L^2 = a^2 + (b\frac{d}{\rho})^2 L^2 = a^2 + h^2, \, mm^2 \qquad (7.4\text{-}5)$$

where a is a constant of the order of 5 mm, related to instrument error, which
dominates the uncertainty for short baselines, and h is a fractional constant of
the order 10^{-8} that dominates for longer baselines. The constant h is largely a
geometric effect proportional to the ratio of the orbit uncertainty, d, and the
receiver-to-satellite distance, ρ, reduced by a geometrical factor, b, which is of
the order 0.2. If the satellite positions can be determined to 1 m, it is possible to
attain 1 cm accuracy for a line 100 km long. A satellite position error of 10 cm
permits a baseline accuracy of 1 cm for 1 000 km long line. However, orbital
accuracies better than 20 cm are a considerable challenge for the GPS
satellites mainly because of their large area-to-mass ratio, which make radiation
pressure effects on the satellite difficult to model.

7.5   SATELLITE RADAR ALTIMETRY

In satellite altimetry the surface topography is determined by a radar altimeter which measures the time for a transmitted pulse to travel to the surface and back again to the satellite (Figure 7.5-1). The distance between the satellite and the surface, combined with a knowledge of the satellite orbit, is then used to construct a global map of the shape of the surface. Particularly important are observations of spacecraft heights over the ocean because this surface approximates to an equipotential surface so that the measurement provides a direct representation of the shape of the ocean, and thereby a good approximation of the geoid. If the satellite carrying the altimeter repeats its track exactly, then changes with time in the sea surface can be obtained.

Uncertainties in the measurement of the seasurface are contributed by surface waves, which distort the reflected altimeter pulse; air molecules and water vapour, which delay the pulse in the *troposphere* (altitude 0-10 km); free electrons in the *ionosphere* (altitude 50-1000km) which further delay the pulse; and the instrument itself. Taken together, these can amount to an uncertainty of 10-20 cm under typical conditions. Fortunately, recent technology developments allow these sources of error to be reduced to acceptable levels. Ionospheric effects can be measured by a dual-frequency altimeter. Watervapour is measured by multichannel radiometers, and wave height is independently measured by the altimeter itself. Only uncertainties in the satellite orbit and in the ability to separate the instantaneous seasurface from the geoid remain as important error sources. Orbits can be calculated with present techniques to an uncertainty of about 30 cm in the height of the satellite.

Satellite altimeters flown on Skylab, Geos-3 and Seasat have shown the promise of this technique. Observations from Geos-3 and Seasat have been combined to map the marine geoid with an accuracy of 1 m over a grid with roughly 100 to 300 km spacing. The three-month Seasat observations have been used to map variable ocean currents. In the future, the TOPEX and ERS-1 missions, both of which are equipped with precision altimeters, will provide new information on the geoid.

7.6   MEASUREMENT ERRORS

Errors arise from various sources: instabilities of the satellite and receiver clocks, orbit errors, delays in the signals caused by uncertainties in propagation velocity through the earth's ionosphere and troposphere, and instrumental errors such as electronic path delays. These errors or *biases* may be removed, or their effect reduced, by post-calibration procedures, by differencing the observations or, as we saw in the previous example, by taking more measurements than are necessary.
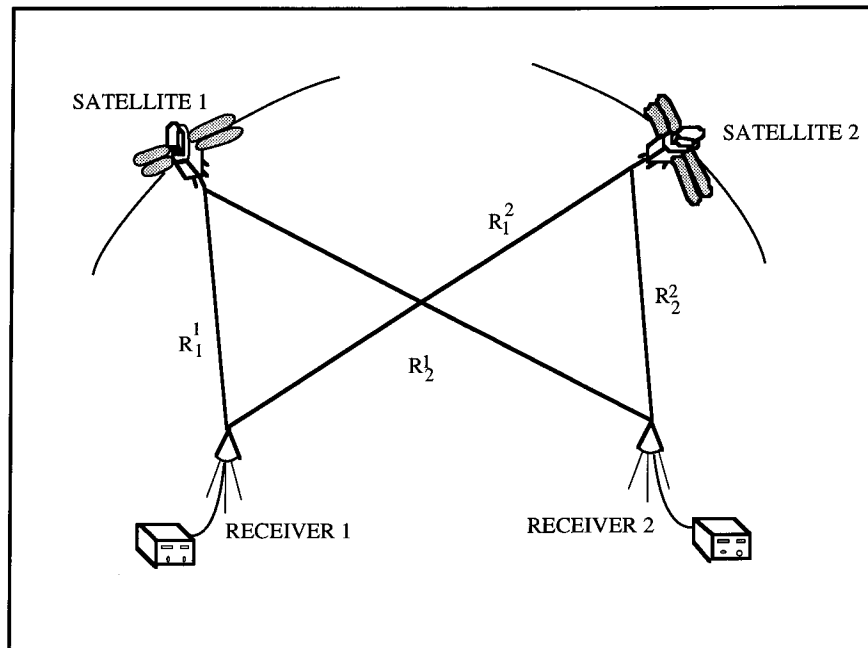
Figure 7.6-1

Basic geometry of differential GPS measurements. The satellites are observed simultaneously.

Example 7.6-1

Figure 7.6-1 depicts the basic geometry of differential GPS measurements. If the ranges between the satellites and receivers are measured simultaneously, *show* that the measurement biases introduced by errors in the satellite clocks are removed by differencing the pseudoranges between satellites (*single differencing*). Hence *show* that the biases introduced by errors in the receiver clocks are removed by differencing the single differences (*double differencing*).

Solution

We rewrite (7.4-2) as follows:

$$R_i^k = \rho_i^k + c\Delta t_i - c\Delta t^k$$

where the subscript, i, denotes the receiver and the superscript, k, denotes the satellite.

The pseudoranges for receiver 1 (i = 1, k = 1,2) are:

$$R_1^1 = \rho_1^1 + c\Delta t_1 - c\Delta t^1 \tag{1}$$

$$R_1^2 = \rho_1^2 + c\Delta t_1 - c\Delta t^2 \tag{2}$$

The pseudoranges for receiver 2 (i = 2, k = 1,2) are:

$$R_2^1 = \rho_2^1 + c\Delta t_2 - c\Delta t^1 \tag{3}$$

$$R_2^2 = \rho_2^2 + c\Delta t_2 - c\Delta t^2 \tag{4}$$

Subtracting (2) from (1) and (4) from (3) ie. differencing between satellites gives:

$$R_1^1 - R_1^2 = \nabla_1^{12} = \rho_1^1 - \rho_1^2 - c(\Delta t^1 + \Delta t^2) \tag{5}$$

$$R_2^1 - R_2^2 = \nabla_2^{12} = \rho_2^1 - \rho_2^2 - c(\Delta t^1 + \Delta t^2) \tag{6}$$

This eliminates the receiver clock biases, Δt_1 and Δt_2.

Subtracting (6) from (5) ie. differencing between satellites and receivers:

$$\nabla_1^{12} - \nabla_2^{12} = \Delta\nabla_{12}^{12} = \rho_1^1 - \rho_1^2 - \rho_2^1 + \rho_2^2$$

This eliminates the satellite clock biases, Δt^1 and Δt^2.

Note

Each time we difference we lose information, in this instance the geocentric location of the receivers ie. the method gives only relative position. Also, the number of measurements is reduced, from four to one here. As a result, positions obtained from double differences are not as well determined as positions obtained from the range measurements.

7.7 DETERMINING GEODETIC PARAMETERS

The basic geometry for estimating geodetic parameters from earth-based satellite observations is illustrated in Figure 7.4-3. The observation vector, $\rho(t)$, of the satellite position relates to the geocentric positions, $r_\Upsilon(t)$, of the spacecraft and, $R_\Upsilon(t)$, of the station according to:

$$r_\Upsilon(t) = R_\Upsilon(t) + \rho(t) \tag{7.7-1}$$

where all three vectors are defined in the inertial reference frame $(X,Y,Z)_\Upsilon$. In general, the vector ρ is only partially observed, for example, its magnitude, ρ, in SLR. The geocentric satellite position vector, $r_\Upsilon(t)$, is a function of the orbital
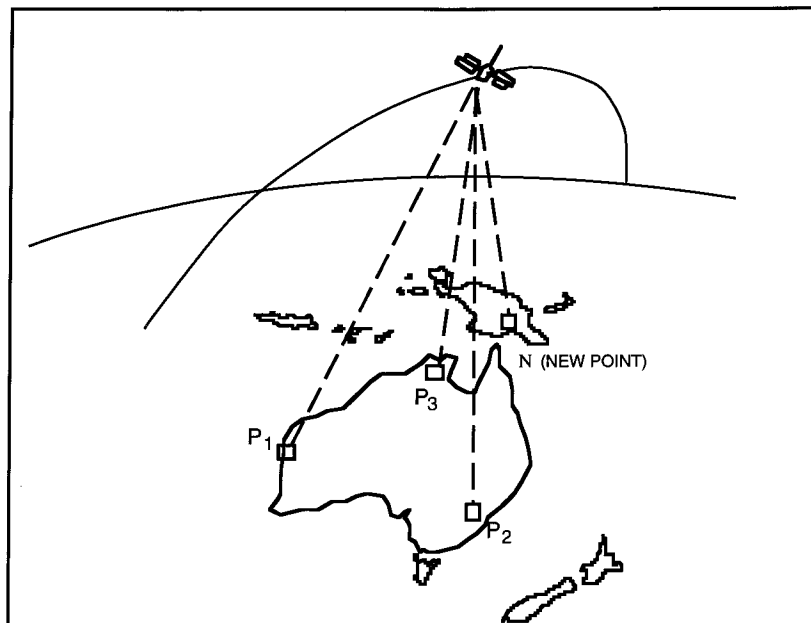
Figure 7.7-1

Geometric method of satellite geodesy. The satellite simply is an elevated and distant target. Observations are made simultaneously at two or more sites.
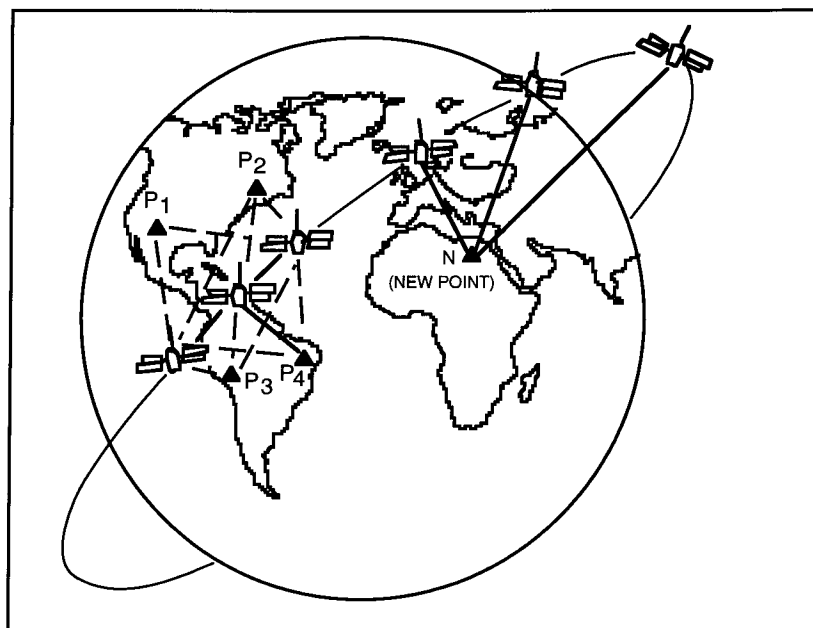
Figure 7.7-2

Dynamic method of satellite geodesy. A tracking network determines the satellite orbit. New points are then located by tracking the satellite over a period of time.

elements at some reference epoch, t_o, and the changes in these elements due to the effect of the various perturbing forces in the interval, $t - t_o$.

The geocentric position vector, $\mathbf{R}_\Upsilon(t)$, of the tracking station transforms to the earth-fixed coordinates, $\mathbf{R}_\oplus(t)$, according to the relation:

$$\mathbf{R}_\oplus(t) = [\mathbf{W}(t)] \, [\mathbf{T}(t)] \, [\mathbf{N}(t)] \, [\mathbf{P}(t)] \, \mathbf{R}_\Upsilon(t) \tag{7.7-2}$$

where $\mathbf{R}_\oplus$ must, in general, be considered as unknown and, at the present level of accuracy of tracking data, be taken as time dependent due to tidal displacements and crustal deformation of the earth. The parameters defining polar motion, $(\mathbf{W})$, and changes in length-of-day, $(\mathbf{T})$, vary in an irregular and unpredictable manner and are also unknown time-varying quantities unless they have been measured by independent methods, while those defining precession, $(\mathbf{P})$, and nutation, $(\mathbf{N})$, can probably be assumed known with sufficient accuracy for most applications.

The above equations describe a complicated nonlinear relation between the observed quantities and astronomical, geodetic, and geophysical parameters. A solution is obtained by observing satellites or radiosources from a large number of stations which are geographically well distributed around the world. Two basic computation methods are used in satellite geodesy: (1) the geometric method, and (2) the dynamic method. In the *geometric method*, the satellite simply becomes an elevated and distant target. Observations are made simultaneously at two or more stations (Figure 7.7-1). The unknown positions of the satellite are eliminated in the observation equations by differencing the measurements leaving only the site positions and measurement biases. Thus the accuracy which is achievable is limited mainly by these biases. Obviously, the method produces only relative positions. If geocentric coordinates of one or more network stations are available, the geometric system can be centred at the geocentre. In the *dynamic method*, we first use a tracking network of sites to determine the satellite orbital parameters (Figure 7.7-2). Since the position of the spacecraft is now known we can locate new points by tracking the satellite over a period of time. Dynamic methods produce absolute positions and the accuracy which is achievable is limited mainly by the accuracy of the computed orbit.

Example 7.7-1

Show that a satellite range bias contaminates mainly the vertical coordinate of position.

Solution

Because atmospheric refraction of the signal is difficult to model at low elevation angles, we usually begin tracking the satellite from the time it rises above an elevation angle of about 15° to the time it is within about 15° of setting below the
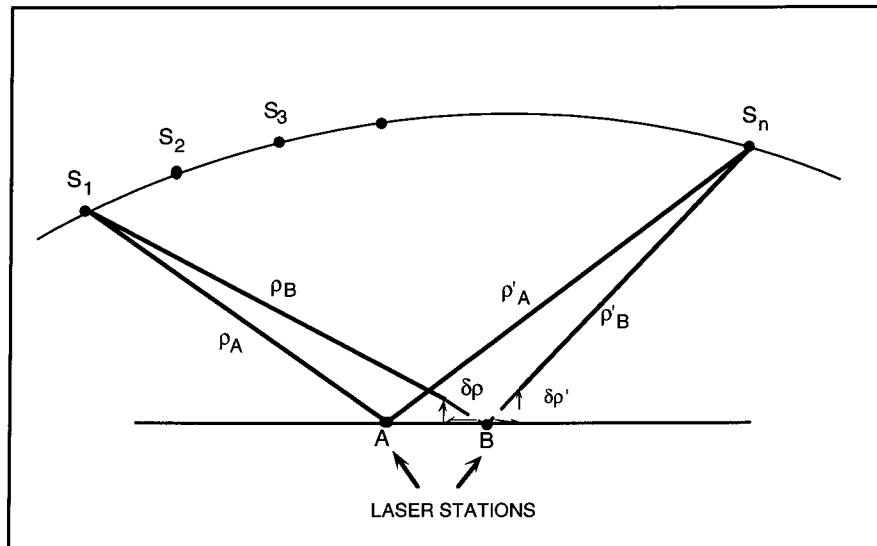
Figure 7.7-3
Effect of a range measurement bias on position determination.

horizon. This tracking interval is known as a satellite *pass*. Figure 7.7-3 shows the effect of a range bias on both the horizontal and vertical coordinates of position. Clearly, the effect of the bias on the horizontal coordinates of position averages out over the duration of the pass. However, the vertical coordinate is never correct. We conclude that, in the presence of measurement biases, the vertical coordinate of position is less well determined than the horizontal coordinates.

Altimeter observations of the height of a satellite above the seasurface, h(t), can be expressed by an equation similar to (7.7-1); however, $R_\Upsilon(t)$, now refers to distance of the seasurface (or geoid) from the geocentre (Figure 7.5-1). Thus:

$$h(t) = r_\Upsilon(t) - R(t) \tag{7.7-3}$$

VLBI determines the rotational motion of the baselines relative to the radio sources (inertial space) and, assuming that these baselines are rigidly attached to the earth, of the polar motion, changes in length-of-day, precession and nutation. In addition, the orientation of the baselines is also changed by differential tidal displacements of the telescopes and by tectonic deformation of the crust. Some of these time-dependent orientation changes can be determined if a number of baselines are observed; for instance, rotational motions will not deform the geometric figures formed by the baselines, nor will they change the baseline lengths. Tectonic and tidal displacements, on the other hand, will generally deform the baseline figures and introduce changes in length, but on different time scales.
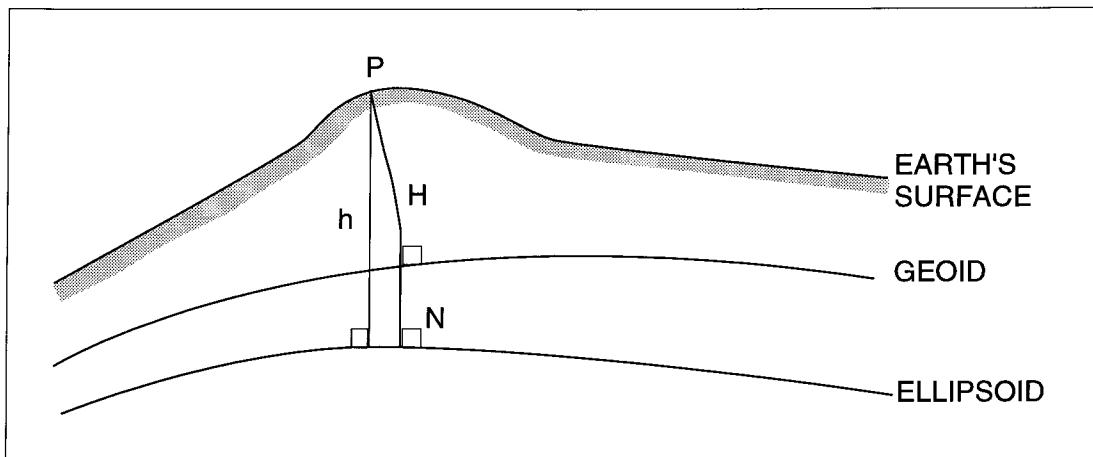
Figure 7.8-1

Relationship between the geoidal undulation, N, ellipsoidal height, h, and orthometric height, H.

7.8   SPACE GEODETIC METHODS AND ORTHOMETRIC HEIGHT

Space geodetic methods give Cartesian coordinates in the ITRF. These may be transformed to geodetic coordinates by defining a horizontal datum with a known relationship to the ITRF. WGS84 is such a datum. If the orthometric height, H, is required, the geoidal undulation, N, must be determined. Figure 7.8-1 shows the geometry of the problem. Since the curvature of the plumbline is very small and for all practical purposes equals the straight line distance, we have:

$$H = h - N \qquad\qquad (7.8-1)$$

The geoidal undulation may be determined by astrogeodetic methods, geometric methods, gravimetric methods, from models for the earth's gravity field, or by combining earth gravity models with local gravity. In the astrogeodetic method positions determined both astronomically and geodetically are compared. The astronomic coordinates refer to the geoid while the geodetic positions refer to the reference ellipsoid. The difference gives the deflection of the vertical which is used to compute the change in N between sites. This can be done with an accuracy of about 2 m. However, the method is difficult to undertake and therefore expensive. As such, it is rarely used today.

The geometric method has been successfully applied in small areas where both the orthometric and ellipsoidal heights are known for a number of well distributed control points. Equation (7.8-1) is used to obtain N at these sites. At other points, where only the ellipsoidal height is available, N is determined by interpolation eg. if enough control stations are available in the area a plane or surface may be fitted to the data. Centimetre-level precision orthometric heights

ACT GEOID
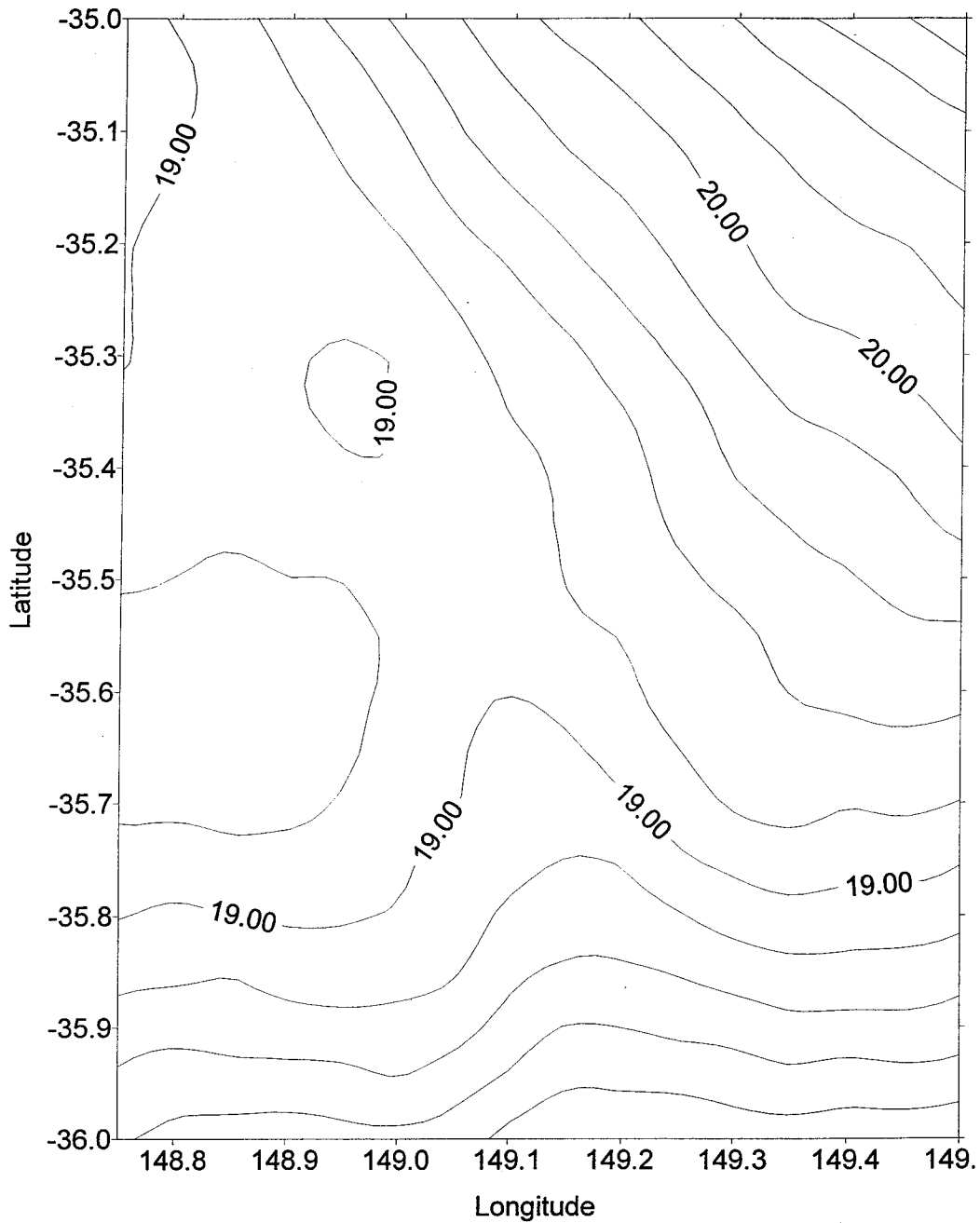
Figure 7.8-2
*The geoid for the Australian Capital Territory from a global geopotential models,
local gravity measurements and terrain data. The contour interval is 0.2 m.
(Courtesy Jim Steed, AUSLIG).*

can be obtained by this method. A proviso is that the distribution of control sites must be reasonably dense (say 10 sites for an area 15 km x 20 km).

In Section 2.9 we describe how N is calculated using Stokes' theorem. The gravity anomalies required for the integration are obtained from terrestrial gravity measurements and satellite altimetry. The earth is divided into a number of zones which are further subdivided into compartments. Each compartment's contribution to N at the site of interest is evaluated and then summed. The accuracy of the absolute determination is about 1-2 m. Changes in N between sites are considerably more accurate, on the order of 2-3 parts in 10^6.

Earlier on we saw that the geoid may be built up from gravity harmonics. The harmonics form what is called a geopotential model. Many such models exist with the better known ones being produced at Ohio State University eg. OSU86. The coefficients of these models are calculated from satellite orbit perturbation analysis, terrestrial gravity measurements and satellite altimetry. High-order representations can furnish N with an accuracy of a few metres for any part of the earth. Higher accuracy, about 1-2 parts in 10^6 for ΔN, may achieved over smaller regions by combining the global models with local gravity and terrain data. Figure 7.8-2 shows such a high-accuracy geoidal model for the Canberra region of Australia.

REFERENCES

BILHAM, R., 1991. Earthquakes and sea level: space and terrestrial metrology on a changing planet, *Reviews of Geophysics*, **29**, 1-29.

BRUNNER, F. & WELSCH, W.M., 1993. Effects of the troposphere on GPS measurements, *GPS World*, January, 42-51.

CARTER, W.E. & ROBERTSON, D.S., 1986. Studying the earth by very long baseline interferometry, *Scientific American*, **11**, 46-54.

DEGNAN, J.J. & PAVLIS, E.C., 1994. Laser ranging to GPS satellites with centimeter accuracy, *GPS World*, September, 62-70.

DIXON, T.H., 1991. An introduction to the Global Positioning System and some geological applications, *Reviews of Geophysics*, **29**, 249-276.

HAJELA, D., 1990. Obtaining centimetre-precision heights by GPS observations over small areas, *GPS World*, January/February, 55-59.

KING, R.W., MASTERS, E.G., RIZOS, C., STOLZ, A. & COLLINS, J., 1987. *Surveying with the Global Positioning System*, F. Dümmler, 128pp.

KLOBUCHAR, J.A., 1991. Ionospheric effects on GPS, *GPS World*, April, 48-51.

LAMBECK, K., 1988. *Geophysical Geodesy*, Clarendon Press, Oxford, 718pp.

LANGLEY, R., 1991. The mathematics of GPS, *GPS World*, July/August, 45-50.

LANGLEY, R., 1993. The GPS observables, *GPS World*, April, 52-59.

MANNING, J. & STEED, J.B., 1994. Geodetic coordinate systems in Australia, *Preview*, **51**, 19-26.

SCHWARZ, K.P. & SIDERIS, M.G., 1993. Heights and GPS, *GPS World*, February, 50-56.

SEEBER, G., 1993. *Satellite Geodesy*, Walter de Gruyter, Berlin, 531pp.

STEED, J.S., 1990. A practical approach to transformation between commonly used reference systems, *The Australian Surveyor*, **35**, 248-264.

STEED, J.B., 1995. The geocentric datum of Australia, *Surveying World*, **4** (1), 14-17.

WANNIGER, L., 1993. Effects of the equatorial ionosphere on GPS, *GPS World*, July, 48-54.

YUNCK, T.P., 1992. A new chapter in precise orbit determination, *GPS World*, October, 56-61.

EXERCISES

7.1 Give the range equation. Explain what it means.

7.2 Explain why the vertical coordinate of position is less well determined than the horizontal coordinates in satellite positioning.

7.3 GPS range observables determined from phase measurements on the carrier wave are ambiguous. What does this mean?

7.4 Distinguish between the geometric and dynamic methods of satellite positioning. What are the major error sources of each method.

7.5 The equation relating satellite position, observer position and satellite measurements is of the form:

$\rho(t) = r(t) - R(t)$

where,

$r(t)$ = position vector of the satellite at time t

$R(t)$ = position vector of the observer at time t

$\rho(t)$ = vector of observations defining the location of the satellite relative to the observer

(i) Give four geodetic parameters implicit in $R(t)$.

(ii) Give two geodetic parameters implicit in $r(t)$.

(iii) Name two methods whereby measurement biases may be eliminated from $\rho(t)$. Why must the satellites be observed simultaneously for one of these to work?

7.6   Explain how SLR works.

7.7   Describe the LAGEOS-type satellites. Why are they spherical and so dense?

7.8   Name two Australian SLR tracking sites.

7.9   Explain how VLBI works.

7.10   Name three Australian VLBI observatories equipped for geodesy.

7.11   Explain how GPS works.

7.12   What is a pseudorange?

7.13   In what way are SLR, VLBI and GPS complementary?

7.14   Distinguish between one-way and two-way ranging systems.

7.15   Describe the ionosphere and explain how ionospheric effects on geodetic observables are reduced in practice.

7.16   Describe the troposphere and explain how tropospheric effects on geodetic observables are reduced in practice.

7.17   Show how geoid height is determined from satellite altimetry.

7.18   Name two satellites used for altimetry.

7.19   In SLR and GPS the observer's position is determined by the intersection of three spheres of radius equal to the measured range and centred on the satellite. Explain how this problem is solved in practice. Why might several iterations be required for the solution to converge?

7.20   Using diagrams to illustrate your answer explain how geodetic parameters such as station position, gravity harmonics, polar motion, precession, nutation, crustal motion and sealevel variations are determined from space geodetic measurements.

7.21   Show that the difference in arrival time, t, of the signal from the radio source at the antennas of an interferometer is given by:

$$\tau = (\Delta x \cos\delta \cos\alpha - \Delta y \cos\delta \sin\alpha + \Delta z \sin\delta)/c$$

where,

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \text{the vector separation of the antennas}$$

α = right ascension of the source

δ = declination of the source

c = velocity of light in vacuo

Hence show that Δz is poorly determined from the variation of t with time.

Index

Publications from the

SCHOOL OF SURVEYING AND SPATIAL INFORMATION SYSTEMS
(formerly: SCHOOL OF GEOMATIC ENGINEERING)
THE UNIVERSITY OF NEW SOUTH WALES
ABN 57 195 873 179

NOTE ON MAIL ORDERS: ALL ORDERS MUST BE PREPAID.
NO REQUESTS FOR RETURNS ARE ACCEPTED.
CREDIT CARDS ARE ACCEPTED. CONDITIONS APPLY. SEE ORDER FORM FOR DETAILS.
WHEN PAYING BY *CHEQUE*, THE CHEQUES MUST BE PAYABLE TO 'SCHOOL OF SURVEYING & SIS, UNSW',
IN AUSTRALIAN DOLLARS AND DRAWN ON AN AUSTRALIAN BANK.

MONOGRAPHS

Australian prices include GST #. Overseas prices include delivery by UNSW's air-lifted mail service (~2-4 weeks to Europe and North America) #. Rates for air mail through Australia Post on application.

**For all** (local and overseas) **orders, add $11.00 for processing and handling** (once per order).

(Prices effective May 2009)

		Price # Australia (incl. GST)	**Price #** Overseas
M1.	R. S. Mather, "The Theory and Geodetic Use of some Common Projections", (3rd edition), 125 pp, reprint 1998, ISBN 0-85839-007-8	$ 16.50	$ 15.00
M2.	R. S. Mather, "The Analysis of the Earth's Gravity Field", 172 pp, 1971 (no ISBN)	$ 8.80	$ 8.00
M8.	A. H. W. Kearsley, "Geodetic Surveying", 96 pp, revised reprint 1988, ISBN 0-85839-036-1	$ 13.20	$ 12.00
M11.	W. F. Caspary, "Concepts of Network and Deformation Analysis", 183 pp, 3rd corrected impression, 2000, ISBN 0-85839-044-2	$ 27.50	$ 25.00
M12.	F. K. Brunner, "Atmospheric Effects on Geodetic Space Measurements", 110 pp, 1988, ISBN 0-85839-048-5	$ 17.60	$ 16.00
M13.	B. R. Harvey, "Practical Least Squares and Statistics for Surveyors", (3rd revised and extended edition), 341 pp, 2006, ISBN 0-7334-2339-6	$ 38.50	$ 35.00
M14.	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management", 269 pp, 1991, ISBN 0-85839-061-2	$ 22.00	$ 20.00
M15/1	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol. 1", 295 pp, 1993, ISBN 0-85939-064-7	$ 33.00	$ 30.00
M15/2	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol. 2", 376 pp, 1993, ISBN 0-85839-065-5	$ 33.00	$ 30.00
M16.	A. Stolz, "An Introduction to Geodesy", 2nd extended edition, 148 pp, 2001, Reprint 2009, ISBN 0-7334-1736-1	$ 24.20	$ 22.00
M18.	J. M. Rüeger, "Electronic Surveying Instruments – A Review of Principles, Problems and Procedures", 1st ed., 166 pp, 2003, ISBN 0-7334-2083-4	$ 25.30	$ 23.00

UNISURV REPORTS - S SERIES

Australian Prices *:	S8 - S20	$11.00 #	S29 onwards	$33.00 #
Overseas Prices **:	S8 - S20	$10.00 #	S29 onwards	$30.00 #

S17.   C. Rizos, "The Role of the Gravity Field in Sea Surface Topography Studies", Unisurv S17, 299 pp, 1980, ISBN 0-85839-029-9

S19.   R. Coleman, "A Geodetic Basis for Recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S19, 332 pp, 1981, ISBN 0-85839-029-9

S36.   A. R. Marshall, "Network Design and Optimisation in Close Range Photogrammetry", Unisurv S36, 249 pp, 1989, ISBN 0-85839-054-X

S37.   W. Jaroondhampinij, "A Model of Computerised Parcel-Based Land Information System for the Department of Lands, Thailand", Unisurv S37, 281 pp, 1989, ISBN 0-85839-055-8

S39.   C. Bosloper, "Multipath and GPS Short Periodic Components of the Time Variation of the Differential Dispersive Delay", Unisurv S39, 214 pp, 1990, ISBN 0-85839-057-4

S40.   J. M. Nolan, "Development of a Navigational System Utilising the Global Positioning System in a Real Time, Differential Mode", Unisurv S40, 163 pp, 1990, ISBN 0-85839-058-2

S41.   R. T. Macleod, "The Resolution of Mean Sea Level Anomalies along the NSW Coastline Using the Global Positioning System", Unisurv S41, 278 pp, 1990, ISBN 0-85839-060-4

S42.   D. A. Kinlyside, "Densification Surveys in New South Wales - Coping with Distortions", Unisurv S42, 209 pp, 1992, ISBN 0-85839-062-0

S43.   A. H. W. Kearsley (ed.), Z. Ahmad, B. R. Harvey and A. Kasenda, "Contributions to Geoid Evaluations and GPS Heighting", Unisurv S43, 209 pp, 1993, ISBN 0-85839-063-9

S44.   P. Tregoning, "GPS Measurements in the Australian and Indonesian Regions (1989-1993)", Unisurv S44, 134 + xiii pp, 1996, ISBN 0-85839-068-X

S45.   W.-X. Fu, "A Study of GPS and Other Navigation Systems for High Precision Navigation and Attitude Determinations", Unisurv S45, 332 pp, 1996, ISBN 085839-069-8

S46.   P. Morgan et al, "A Zero Order GPS Network for the Australia Region", Unisurv S46, 187 + xii pp, 1996, ISBN 0-85839-070-1

S47.   Y. Huang, "A Digital Photogrammetry System for Industrial Monitoring", Unisurv S47, 145 + xiv pp, 1997, ISBN 0-85839-072-8

S48.   K. Mobbs, "Tectonic Interpretation of the Papua New Guinea Region from Repeat Satellite Measurements", Unisurv S48, 256 + xv pp, 1997, ISBN 0-85839-073-6

S50.   M. D. Subari, "Low-cost GPS Systems for Intermediate Surveying and Mapping Accuracy Applications", Unisurv S50, 179 + xiii pp, 1997, ISBN 0-85839-075-2

S51.   L.-S. Lin, "Real-Time Estimation of Ionospheric Delay Using GPS Measurements", Unisurv S51, 199 + xix pp, 1997, ISBN 0-7334-1664-0

S53.   D. B. Lemon, "The Nature and Management of Positional Relationships within a Local Government Geographic Information System", Unisurv S53, 273 + xvi pp, 1997, ISBN 0-7334-1678-0

S54.   C. Ticehurst, "Development of Models for Monitoring the Urban Environment Using Radar Remote Sensing", Unisurv S54, 282 + xix pp, 1998, ISBN 0-7334-1679-9

S55.   S. S. Boey, "A Model for Establishing the Legal Traceability of GPS Measurements for Cadastral Surveying in Australia", Unisurv S55, 186 + xi pp, 1999, ISBN 0-7334-0685-X

S56.   P. Morgan and M. Pearse, "A First-Order Network for New Zealand", Unisurv S56, 134 + x pp, 1999, ISBN 0-7334-0685-8

S57.   P. N. Tiangco, "A Multi-Parameter Radar Approach to Stand Structure and Forest Biomass Estimation", Unisurv S57, 319 + xxii pp, 2000, ISBN 0-7334-0786-2

S58.   M. A. Syafi'i, "Object–Relational Database Management Systems (ORDBMS) for Managing Marine Spatial Data: ADCP Data Case Study", Unisurv S58, 123 + ix pp, 2000, ISBN 0-7334-1704-3

S59.   X.-Q. Lu, "Strategies for Improving the Determination of Displacements of Sea Surface Temperature Patterns Using Consecutive AVHRR Thermal Images", Unisurv S59, 209 + xiii pp, 2000, ISBN 0-7334-1721-3

S60.   G. Dickson, "GPS-Controlled Photography: The Design, Development and Evaluation of an Operational System Utilising Long-Range Kinematic GPS", Unisurv S60, 417 + x pp, 2000, ISBN 0-7334-1725-6

S61.   J. Wang, "Modelling and Quality Control for Precise GPS and GLONASS Satellite Positioning", Unisurv S61, 171 + x pp, 2001, ISBN 0-7334-1766-3

S62.   Y. Wang, "Knowledge-Based Road Extraction from Aerial Images", Unisurv S62, 178 + xi pp, 2001, ISBN 0-7334-1767-1

S63.   L. Ge, "Development and Testing of Augmentations of Continuously-Operating GPS Networks to Improve their Spatial and Temporal Resolution", Unisurv S63, 230 + xvi pp, 2001, ISBN 0-7334-1841-4

S64.   H.-Y. Chen, "A Study on Real-Time Medium-Range Carrier-Phase-Based GPS Multiple Reference Stations", Unisurv S64, 182 + xxiv pp, 2001, ISBN 0-7334-1842-2

S65.   G. Y. K. Shea, "A Web-Based Approach to the Integration of Diverse Data Sources for GIS", Unisurv S65, 233 + xv pp, 2001, ISBN 0-7334-1845-7

S66.   M. Mirbagheri, "Analysis of Interferometric SAR for Topographic Mapping", Unisurv S66, 135 + xvii pp, 2001, ISBN 0-7334-1856-2

S67.   P. Wang, "Applying Two-Dimensional Kalman Filtering Techniques to Digital Elevation Models for Terrain Surface Modelling", Unisurv S67, 175 + xi pp, 2001, ISBN 0-7334-1857-0

S69.   C. Satirapod, "Improving the GPS Data Processing Algorithm for Precise Static Relative Positioning", Unisurv S69, 131 + viii pp, 2002, ISBN 0-7334-1901-1

S70.   R. Mason, "Developing Australian Spatial Data Policies – Existing Practices and Future Strategies", Unisurv S70, 258 + xv pp, 2002, ISBN 0-7334-1941-0

S71.   C. Ogaja, "A Framework in Support of Structural Monitoring by Real Time Kinematic GPS and Multisensor Data", Unisurv S71, 191 + xiii pp, 2002, ISBN 0-7334-1958-5

S72.   L. Dai, "Augmentation of GPS with GLONASS and Pseudolite Signals for Carrier Phase Based Kinematic Positioning", Unisurv S72, 188 + viii pp, 2002, ISBN 0-7334-1975-5

S73.   C. Roberts, "A continuous Low-Cost GPS-Based Volcano Deformation Monitoring System in Indonesia", Unisurv S73, 271 + xvi pp, 2002, ISBN 0-7334-1976-3

S74.   V. Janssen, "A Mixed-Mode GPS Network Processing Approach for Volcano Deformation Monitoring", Unisurv S74, 199 + viii pp, 2003, ISBN 0-7334-2059-1

S75.   Y. H. Lu, "Automatic Building Extraction for 3D Terrain Reconstruction Using Image Interpretation Techniques", Unisurv S75, 175 + xii pp, 2004, ISBN 0-7334-2141-5

S76.   Y. K. Lee, "Integration of GPS/Pseudolites/INS for High Precision Kinematic Positioning and Navigation", Unisurv S76, 200 + xiv pp, 2004, ISBN 0-7334-2149-0

Note: From 1 January 2005, PhD and MEng theses are nor longer published in this series. See   www.gmat.unsw.edu.au   for information on post 2004 theses.

School of Surveying and Spatial Information Systems
The University of New South Wales
ABN 57 195 873 179

ORDER FORM

To order, mail or fax this form to: Publications Officer, School of Surveying and Spatial Information Systems, The University of New South Wales, UNSW SYDNEY NSW 2052, AUSTRALIA   Fax: +61 - 2 - 9313 - 7493

Name and postal address of ordering person/organisation:

_____

_____

_____

_____

_____

Email : _____   Fax : _____

Day time phone number : _____   Date : _____

Note: All orders must be prepaid.   No requests for returns are accepted.

I/We order herewith:

_____	$	_____
_____	$	_____
_____	$	_____
Postage and Handling (from 1 July 2004)	$	**11.00**
Total	$	_____

Payment Details:

☐   A cheque, drawn on an Australian Bank, in Australian Dollars and payable to 'School of Surveying & SIS, UNSW' is enclosed

☐   Please debit my credit card: ☐ Mastercard   ☐ Visa

Name on Card: _____   Signature: _____

Card No: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐   Expiry Date : _____