UNIVERSIDADE FEDERAL DA BAHIA
CENTRO DE PESQUISA EM GEOFÍSICA E GEOLOGIA

MOTIONS, GRAVITY FIELD AND FIGURE OF THE EARTH

Oldrich Novotný

Lecture notes for post-graduate studies

Instituto de Física
Instituto de Geociências

Salvador, Bahia, 1998

Contents

Preface

Physics is traditionally divided into smaller branches such as mechanics, electricity and magnetism, thermodynamics, atomic physics and others. Mechanics is usually further divided into two main parts:

1) mechanics of the particle, of a system of particles, and of the rigid body;
2) continuum mechanics.

Based on this division of physics, geophysics is also divided into analogous branches; for details refer to the next chapter.

The present lecture notes represent an introduction to those parts of mechanics of the Earth which, in physics, correspond to particle mechanics and rigid body mechanics. In particular, we shall deal with the motions of the Earth, with its gravity field and figure.

Most of the problems discussed in these lecture notes are the subject of one of the geophysical disciplines, called gravimetry. However, gravimetry does not include all the topics. Although gravimetry studies the Earth's gravity field in great detail, the other problems are usually studied elsewhere. Hence, the motions of the Earth or satellite methods of investigating the gravitational field are considered to be topics of astronomy rather than of geophysics. Problems of the figure of the Earth also traditionally are a part of geodesy, although they have much in common with gravimetry.

Consequently, the present lecture notes partly have an interdisciplinary character, containing chapters from astronomy, geophysics and geodesy. Contrary to specialized textbooks and monographs on these problems, our explanation is more elementary. The specialized books often assume that the reader is acquainted with extensive chapters on coordinate systems, celestial mechanics, or the potential theory. We have tried to shorten these auxiliary chapters and to start immediately with the mechanics of the Earth. It is only assumed that the reader has passed the basic courses of mathematics and physics which are usual at the faculties of natural or technical sciences. More advanced passages of mathematics and physics which are needed here, such as Legendre polynomials or mechanics in non-inertial reference frames, are explained in separate chapters. Some chapters of the lecture notes are relatively independent, so that they can also be used in other lectures.

For many years, I have read a course of gravimetry for undergraduate students of geophysics at the Faculty of Mathematics and Physics of the Charles University in Prague, Czech Republic. I have used the experience gained from these lectures here, but the present text has mostly been extended substantially, and several new chapters have been added.

These lecture notes have been written for the purposes of the post-graduate studies in geophysics, organized by the Universidade Federal da Bahia, Salvador, Brazil.

I would like to thank the Centro de Pesquisa em Geofísica e Geologia, Instituto de Física, and the Instituto de Geosciências for their support in

preparing this text. I wish to express my thanks especially to the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for providing me with the fellowship which made my stay at the Universidade Federal da Bahia possible. I would also like to express my gratitude to the students and professors whose advice and comments contributed to improving the text, in particular to Prof. Roberto Max de Argollo and Prof. Elpídio A. Jucá. I am especially obliged to Mr. Armando Guedes Vicentini for his initiative in preparing the computer version of a substantial part of the notes, and to Mr. Joaquim Bonfim Lago for his work in the technical preparation of the text and figures. I also thank RNDr. Jaroslav Tauer, CSc for the language revision of the text with an understanding of the subject, which made the text more comprehensible in many places. My thanks are also due to my wife, Mrs. Šárka Novotná, for the technical preparation of several of the last chapters. I shall also be grateful to every reader for any critical comments and remarks concerning these notes.

Salvador, 1998

Oldrich Novotný

Introduction

Geophysics and Its Division

During the long historical development of Earth sciences, several independent scientific disciplines have been constituted. The main of these Earth sciences are: geodesy, geology, geophysics, geochemistry and geography (Fig 0.1).

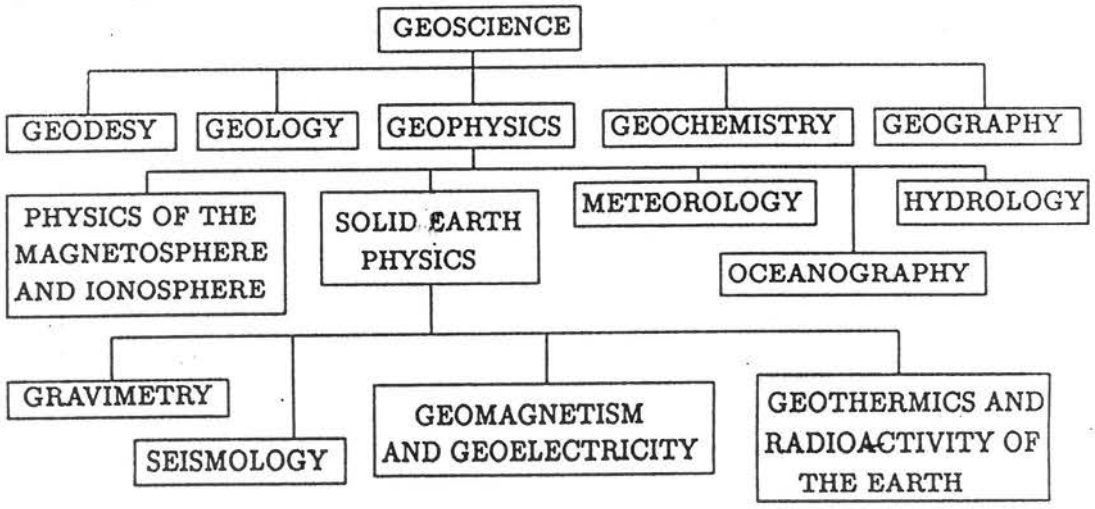

Figure 0.1: Earth sciences and branches of geophysics.

Geophysics (physics of the Earth) is a branch of physics which studies the phenomena and processes occurring in the Earth and in its immediate vicinity by means of physical methods. It also studies the effects which other celestial bodies, in particular the Sun and the Moon, have on the Earth.

From the point of view of this broad definition, geophysics includes physics of the solid part of the Earth and of its liquid and gaseous envelopes. In this case we speak of geophysics in a broader sense. However, meteorology, oceanography and hydrology are usually separated from geophysics as independent scientific disciplines with specific problems and methods of investigation. We then speak of geophysics in a narrower sense or of solid Earth physics. Attention should be paid to a certain terminological inconsistency, when solid Earth physics is understood to be the physics of the whole interior of the Earth including its liquid parts, especially the liquid outer core.

Solid Earth physics is further subdivided into smaller disciplines after certain analogies with physics. The main of these disciplines are (Båth, 1979):

1. Gravimetry: measurement of gravity and its interpretation;
2. Seismology: study of earthquakes and of the propagation of seismic waves;
3. Geomagnetism and geoelectricity: electromagnetic field of the Earth;

4. Geothermics and radioactivity of the Earth: the thermal conditions and heat sources in the Earth's interior.

In addition to these disciplines, some others, which have a rather interdisciplinary character, also belong to geophysics. On the boundary between geodesy and geophysics, there are physical geodesy and the study of recent motions. Between geology and geophysics, there are tectonophysics, volcanology and geochronology (determination of the age of rocks and geological processes). Of a common interest to astronomy and geophysics are the motions of the Earth, satellite gravimetry, planetary physics and cosmogony (study of the origin of the Solar System).

In the same way as physics, geophysics can also be divided using other criteria, e.g., into theoretical and experimental geophysiscs, or into pure and applied geophysics.

Geodesy and Gravimetry

The prevailing part of these lecture notes will be devoted to study of the *gravity field* of the Earth. By "gravity field" we understand the superposition of the gravitational field of the Earth and of the field of the centrifugal force which is generated by the rotation of the Earth about its axis. Much attention will be paid to the form of the equipotential surfaces of the gravity potential. We shall see that the shape of these surfaces is closely related to the figure of the Earth. As a result, many problems of geodesy and gravity blend together. We can say that accurate geodetic measurements must be accompanied by gravimetric measurements, and vice versa. Let us illustrate this on several simple examples (Grushinskii, 1976).

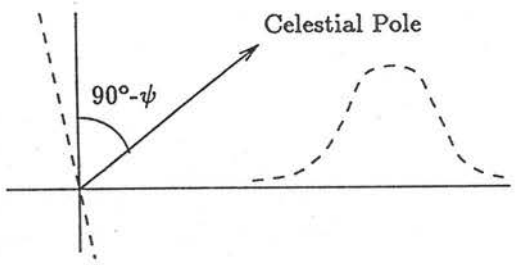

Figure 0.2: The influence of anomalous masses on the astronomical measurement of geographic latitude ψ.

The measurement of the geographic latitude requires the angle between the plumb line and the direction toward the celestial pole (in the Northern Hemisphere in the neighbourhood of Polaris) to be measured. The geographic latitude, ψ, is then the complement of this angle to 90° (Fig. 0.2). However, if there is a pronounced concentration of masses (a hill, mountain range) in the neighbourhood of the point of observation, these masses cause a deflection of the plumb line (see the dashed lines in Fig. 0.2 denoting a hill and the deflected

plumb line). As a result, a deviated value of the geographic latitude will be obtained. Although the effect of a hill on the deflection of the plumb line can be approximately estimated, we know nothing of the distribution of the masses below the hill. Therefore, to determine the deflections of the plumb line more accurately, more sophisticated methods had to be developed (we shall deal with them below).

The deflections of the plumb line, caused by anomalous masses, also affect the determination of the heights above sea level by means of levelling (Fig. 0.3). The problem is that the levelling instrument is adjusted to the local horizontal plane, but this plane is affected by the presence of anomalous masses.

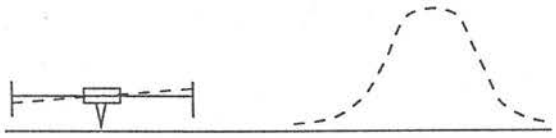

Figure 0.3: The influence of anomalous masses on the determination of the height above sea level by means of levelling.

However, gravity measurements cannot be carried out without accurate geodetic measurements. For example, to construct gravity maps we must know not only the geographic coordinates of the corresponding points, but also the heights above sea level to be able to calculate the gravity reductions.

Methods of determining the figure of the Earth underwent long historical development. We can divide them into three main groups: geometric, gravimetric (physical geodesy) and astronomical methods (satellite geodesy). The oldest of these methods are the geometric; we shall deal with them briefly in the next chapter. The theoretical foundations of some of the gravimetric methods were already laid in the 18th and 19th centuries, but these methods were widely applied in the first half of the 20th century. The geometric methods enable not only the shape, but also the linear dimensions of the Earth to be determined. However, they can be used only on the continents. The gravimetric methods are only able to determine the shape of the Earth, but their great advantage is that they can also be applied on the oceans. The astronomical methods have rarely been used in the past, as the use of the distant Moon was inaccurate for these purposes. However, in connection with man-made satellites, these methods now belong to the most important and most accurate.

Chapter 1

Historical Review of the Study of the Figure of the Earth

1.1 Ancient Mythical Notions

The history of the scientific study of the Earth began with the idea of a spherical Earth. As this concept does not follow from everyday experience, it underwent a long historical development (Fischer, 1975).

Some indications of the oldest notions about the Earth can already be found in ancient myths. Of course, we should not try to infer any profound conclusions from them, as the questions which we are interested in were not formulated in these myths yet.

The oldest preserved written record is a Sumerian-Babylonian epic, Gilgamesh, which is 5 000 years old. Gilgamesh is the name of a man who cannot accept the death of his closest friend and the prospect of his own death some day. He, therefore, decided to look for the secret of immortality. He knew that he had immortal ancestors, Utnapishtim and his wife, who lived somewhere far away. Gilgamesh decided to visit them and to learn the secret of immortality from them. During his journey, full of dangers, he met the immortal gods, who were not in any heaven, but right there among the human mortals. Also the immortal couple did not live somewhere at the end of the world, but at a forbidden place, which was inaccessible because of poisonous waters all around. Gilgamesh managed to overcome these obstacles and at last he returned happily.

From the geodetic point of view we could consider this world as being horizontally infinite, i.e. as if two-dimensional. A more complete notion of the third dimension, the "above" appears later in connection with the formulation of the concept of the sky as a hemisphere studded with stars.

For a long time, it was thought that the Earth had the form of a flat disk which supports the hemispherical sky. This notion was accepted, e.g., by Homer in the 9th century B.C. and by Ionian philosophers still in the 6th century B.C. However, what supports the disk, was a difficult question to answer. The opinions of this problem gradually changed.

According to one mythical notion, the disk was upheld by four elephants who stood on the back of a big turtle swimming in milk (which supported the swimming turtle). Thales of Miletus (about 624 to about 543 B.C.) thought that the disk rested on water. Anaximenes (about 585 to about 525 B.C.) maintained it was held fixed at an infinite depth by air which was barred from escaping upwards by the size of the disk. This infinite support was in contradiction with the general notion that the invisible nightly paths of the Sun and stars were under the Earth. It was necessary to assume that the Sun, from its setting in the west to its rising in the east, returned around the side of the disk, e.g., along the northern horizon. However, as it is dark at night, it was necessary to adopt another

assumption that there had to be high mountains in the north which hid the Sun's nightly path.

Anaximander (611 to 547 B.C.) drew a bold and far-reaching conclusion from these puzzling problems and from the observation of the circumpolar stars. According to him, the Sun and stars completed circular paths even when not seen. To make this possible, he had to remove the infinite support from under the Earth and keep it floating freely, and he had to postulate an invisible hemisphere for the invisible part of the celestial motions. Anaximander's celestial sphere revolved together with all the stars around an unattached, much smaller Earth suspended at the centre. His Earth, as opposed to the celestial sphere, was not spherical yet, but a low cylinder.

The description of these notions and speculative constructions undoubtedly raises a smile in the present reader. However, we should realize that the manner of thinking and of logical argumentation of the ancient philosophers did not differ much from ours. A substantial difference existed only in the extent of concrete knowledge. (Of course, differences did exist. In adopting new theories we emphasize the role of experiments, whereas in antiquity, logical indisputability was emphasized more). On the contrary, Anaximander's notion of a freely floating Earth, or even the speculations about a spherical Earth, which will be dealt with in the next section, should be admired. We should realize that these speculations preceded by many centuries the famous discovery of Newton's gravitational law, which fundamentally influenced our understanding of the world that surrounds us.

1.2 The Spherical Earth

The concept of a spherical Earth is ascribed to Pythagoras (about 580 to 500 B.C.) and his school. Pythagoras did not arrive at the idea of a spherical Earth on the basis of any observation, but apparently as a result of speculative deliberations about the harmony of the world. As the sphere represents a perfect form, the perfect world should also be constructed of planetary bodies of spherical form.

At the time of Aristotle (384 to 322 B.C.) the spherical shape of the Earth was generally known, but there was no really cogent proof thereof. Aristotle himself claimed that all heavy bodies had a tendency to fall towards the centre of the world, which led to a spherical configuration (again an interesting thought containing a seed of the law of universal gravitation). Moreover, Aristotle pointed out the changing horizon as one proceeded north, and the round shadow of the Earth in lunar eclipses.

The speculations about the harmony of the Earth again seem to be very far from contemporary physical ideas. But this is only a first impression. The idea of harmony was used by Kepler in formulating his laws. The principle of simplicity of the world was emphasized by many philosophers and physicists (Descartes, Einstein and others). Even contemporary physics is full of various speculations about harmony, only the word "harmony" is usually replaced by the word

"symmetry". The existence of various laws of conservation follows from symmetries (invariances in transformations), and the idea of symmetry plays an important role in the physics of elementary particles and elsewhere.

1.3 The Size of the Spherical Earth

The first measurement of the size of the Earth, based on a scientific and sufficiently accurate method, was carried out by Eratosthenes (276 to 195 B.C.) in Egypt. He knew that in Syene (now called Aswan) it was possible to see the image of the Sun in a deep well at noon on the day of the summer solstice. This meant that the zenith distance of the Sun was close to zero at that time, which indicated that Syene was close to the Tropic of Cancer. Using a sun dial in the form of a hemispherical bowl, he found that, at the same time in Alexandria, the Sun's rays and the vertical formed an angle of 1/50 of the full circle; we would now speak of 1/50 of 360°, i.e. of an angle of 7°12', see Fig. 1.1, where it is denoted α.

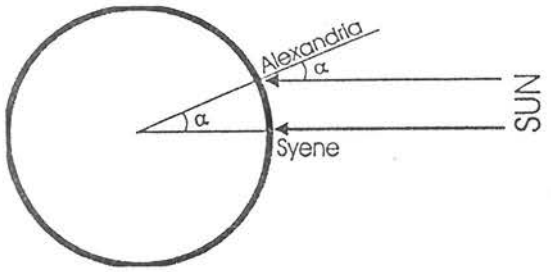

Figure 1.1: Eratosthenes' method of determinig the size of the Earth.

Assuming that both cities lie on the same meridian, it followed from elementary geometry that the length of the arc between Alexandria and Syene must be equal to 1/50 of the circumference of the Earth. He determined the distance between Alexandria and Syene to be 5 000 stadia. Opinions differ whether this value was determined by Eratosthenes from the known travel time of camel caravans, or in a more accurate manner from the Egyptian cadastral maps, already existing at that time. The caravan could, according to Eratosthenes, travel from Alexandria to Syene in 50 days, assuming that the fairly constant speed of the camels was 100 stadia per day. By multiplying 5 000 stadia by fifty he obtained 250 000 stadia for the whole circumference of the Earth. The figure 252 000 is also quoted as Eratosthenes' result. The accurate length of the Egyptian stadium is not known now; estimates usually vary between 148 and 158 m. The circumference of 252 000 stadia then corresponds to the length of the Earth's quadrant within the interval of 9 324 to 9 954 km, which is admirably close to the correct value of 10 000 km. Consequently, Eratosthenes is often considered to be "the father of geodesy". Many later measurements did not yield such good results. (For example, at the time of

Christopher Columbus, the Earth was considered to be much smaller; see Tab. 1.1 below.)

Obviously, Eratosthenes was lucky to have obtained such a good result. It seems that some errors compensated one another. Firstly, Alexandria and Syene do not lie on the same meridian, as Eratosthenes assumed (Syene also does not lie precisely on the Tropic of Cancer at present, but at the time of Eratosthenes' measurement it was very close to it). Furthermore, the determination of the distance of both cities was not accurate, especially if derived from the speed of camel caravans. Nevertheless, the principle of Eratosthenes' method was correct. We now refer to it as the *arc measurement*.

1.4 Triangulation

The oldest methods of determining the size of the Earth were inaccurate especially in determining distances on the Earth's surface. A basic change occurred at the turn of the 16th and 17th centuries, when triangulation was introduced into geodesy as a distance-measuring procedure.

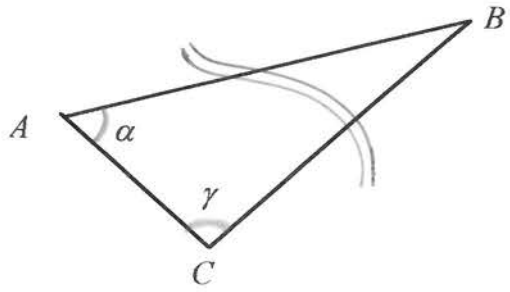

Figure 1.2: Trigonometric determination of distance AB.

Very often we encounter a problem that the distance between two points A and B cannot be measured directly because there is an obstacle between them, such as a river, high hill, dense forest, etc. The well-known trigonometric method is then usually used (Fig. 1.2). We choose an auxiliary point C, measure distance AC and at least two angles in triangle ABC. Distance AB can then be determined by calculation.

If a larger distance is to be determined, a system of triangles can be used. We speak then of *triangulation*; see (Fig. 1.3). Tycho Brahe used the idea of triangulation already in 1589 to determine the distance of a small island from the Danish mainland. (Tycho Brahe was then working in Prague at the court of emperor Rudolph II. He carried out accurate astronomical measurements which were used by Johannes Kepler in deriving his famous laws of planetary motions).

Although the principles of triangulation were known earlier, the decisive step for introducing triangulation into geodesy was the invention of the telescope in 1609. (Galileo Galilei used the telescope for the first astronomical measurements already in 1610. He discovered the largest four moons of Jupiter, i.e. the moons

Io, Europa, Ganymedes and Callisto, which are now referred to as the Galilean satellites of Jupiter). The telescope made it possible to measure even the angles of large triangles accurately.

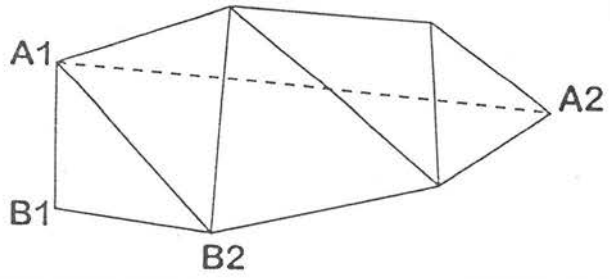

Figure 1.3: Principle of triangulation. Distance A_1A_2 is to be determined. If the length of baseline B_1B_2 and the angles in all triangles are measured, it is possible to compute not only all sides of the triangles, but also distance A_1A_2.

The first to apply triangulation for a better determination of the Earth's circumference using Eratosthenes' method was Willebrord Snellius in 1615. He measured a chain of 33 triangles on a nearly meridional arc in a region on the Lower Rhein in the Netherlands, and combined it with the astronomic latitudes at the endpoints. The resulting radius of the Earth, which was later slightly corrected, differs only by few kilometers from the correct value (Tab. 1.1).

Table 1.1: The radius of the spherical Earth.

Author			Radius, km
Eratosthenes	252 000 stadia,	148 m each	5 936
		158 m each	6 337
Columbus			4 821.4
Snellius	corrected		6 368.7
Picard			6 371.98

Examples are often presented to demonstrate how progress in basic research influences the development of technology. Here we have the opposite case when technical progress (the invention of the telescope) iniciated substantial changes in basic research.

1.5 The Ellipsoidal Earth

The options provided by triangulation inspired the French Government to found the Royal Academy of Sciences (Académie Royale des Sciences) in 1666, for the purpose of improving maps and studying related problems.

One of the first tasks of the Academy consisted in an accurate measurement of the size of the Earth, which was needed for establishing the scale for the latitude-longitude grid. Jean Picard was assigned to this task. In 1669 and 1670

he measured an arc from a point near Paris northward to Amiens, and determined astronomically the latitude difference at the endpoints. Picard's measurement was modern in several aspects. He measured a real base line with the aid of wooden rods, used a telescope in his angle measurements and logarithms in his computations. His chain consisted of 13 large triangles. The significance of this measurement is enhanced by the fact that, when Newton derived his famous gravitational law, he used Picard's value for the size of the Earth (Tab. 1.1).

Between 1683 and 1716 this arc was prolonged northward to Dunkerque (the town known for the heavy fighting during World War II), and southward to the Spanish border. The measurements were carried out by Philippe de Lahire and mainly by the Cassinis (Dominique and Jaques). The measurements were already sufficiently accurate that it was possible to try to determine not only the size of the Earth, but the shape of the meridian as well. The measured arc was divided into two parts, one northward from Paris, the other southward. When they computed the length of a 1° meridional arc independently from both chains, they came to the unexpected result that these lengths were different. This finding was in contradiction with the conception of a spherical Earth. It could only be the result of a deviation of the Earth from spherical shape, or of observational errors. Nevertheless, it was a first indication that, for example, an ellipsoid could approximate the figure of the Earth better than a sphere.

Another problem of these measurements was the fact that the length of the arc, corresponding to 1° of latitude, decreased northward. Expressing lengths in the present units, i.e. in meters, the length of the degree in the northern part of the chain was 111 017 m, whereas in the southern part it was 111 284 m, which was 267 m more (Heiskanen and Vening Meinesz, 1958). At first sight it seems that the decrease of the length of a degree northward can be explained by a decrease of the distances to the Earth's centre, i.e. by the polar flattening of the Earth. This would really be the case if the measured arcs corresponded to the geocentric latitude. However, the situation is more complicated and, actually, the opposite conclusion must be drawn from these measurements. Therefore, let us analyse the problem in greater detail in the next section.

1.6 Geocentric and Geodetic Latitudes

Consider a straight line connecting point P with the Earth's centre O (Fig. 1.4). The geocentric latitude φ of point P is the angle between this straight line and the equatorial plane. Geodetic latitude ψ is the angle between the normal to the Earth's ellipsoid at point P and the equatorial plane. The following relation holds true between these latitudes:

$$\tan\psi = \frac{a^2}{b^2}\tan\varphi \ , \tag{1.1}$$

where a and b are the equatorial and polar semi-axes of the ellipsoid, respectively. Formula (1.1) can easily be derived if we express the equations of the corresponding ellipse in Cartesian coordinates, of its tangent and normal.

In measuring the latitude astronomically, the angle between the vertical and the direction toward the celestial pole is determined (the northern celestial pole lies in the neighbourhood of Polaris). The latitude is then the complement of this angle to 90°. Since the vertical is close to the normal of the ellipsoid, the latitude thus determined represents the geodetic latitude.

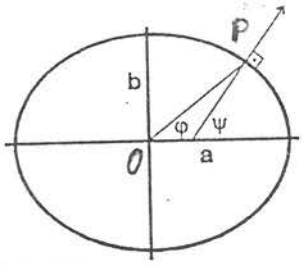

Figure 1.4 The geocentric latitude φ and geodetic latitude ψ of point P. The flattening is exaggerated.

Since the astronomical method of determining latitudes is also used in arc measurements, such latitudes are geodetic. Consequently, a 1° meridional arc is the arc between the corresponding normals to the ellipsoid. This means that in arc measurements we determine the radius of curvature of the ellipsoid, but not the distance from its centre. The result of the arc measurements in France, i.e. the decrease of the length of a degree northward, must then be interpreted as a decrease of the radius of curvature northward. Therefore, according to these measurements, the Earth ellipsoid should be elongated in the polar regions, not flattened.

The geodetic latitude is used in maps. On the other hand, in physics, where spherical coordinates are often used, it is more convenient to work with the geocentric latitude. Since the flattening of the Earth is relatively small, the differences between the geodetic and geocentric latitudes are not very large, the largest being 11.55 arc minutes around latitude 45° $\left(\varphi = 44.9°, \text{i.e. } \psi = 45.1°\right)$.

1.7 Dispute about the Type of Earth Ellipsoid

The arc measurements on the territory of France challenged the millennia-old concept of a spherical Earth. It became evident that it would be necessary to abandon the deep-rooted notion of the geometrically ideal form, a sphere, for the figure of the Earth. The considerations about the harmony or simplicity of the world, which had played such an important role in forming the ancient notions of the Earth's shape, became an obstacle of further progress in the second half of the 17th century. This lesson from the history of geodesy should

also be remembered. It illustrates the limitations of the "principle of simplicity", which has often been used in many branches of science (see comments in Section 1.2).

The results of the arc measurements in France, namely the decrease of the length of a degree northward, indicated that the Earth was elongated at the poles, i.e. egg-shaped, or lemon-shaped (Fig. 1.5). However, this result was in contradiction with the new theories in physics which had been formulated, mainly by Isaac Newton, nearly at the same time .

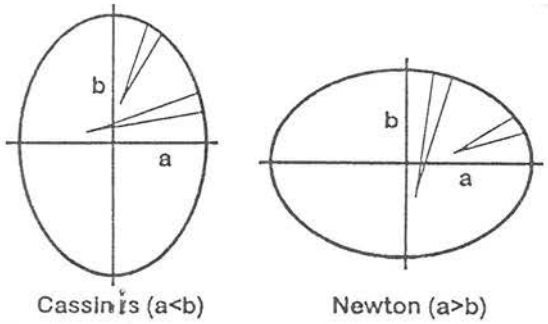

Figure 1.5: Shape of the earth: elongated or flattened at the poles?

Newton published his famous "Philosophiae naturalis principia mathematica" in 1687. Apart from other results, he had formulated the law of universal attraction there. He concluded that, as a result of the superposition of the gravitational and centrifugal forces, the rotating Earth would have to be flatter in the polar regions than near the equator (like a grapefruit, versus the French "lemon").

This situation started an intense controversy between French and English scientists. Frenchmen, particularly the Cassinis, defended their own measurement and were inclined to keep the Earth egg-shaped. However, the English claimed that the Earth must be flattened, as Newton had shown theoretically. This conflict between the French "Earth elongators" and the British "Earth flatteners" aroused attention even outside technical circles.

Although Newton's conclusions were derived from purely theoretical considerations, two important observations supported them: 1) a change of the oscillation period of the pendulum clock between Paris and Guyana (a change of the clock rate); 2) the discovery of the rotation and flattening of Jupiter. Let us add several details related to these observations.

Jean Richer had trouble adjusting his clock on his expedition to Cayenne, Guyana, in South America. Although the clock had been regulated carefully in Paris, it now lost $2\frac{1}{2}$ minutes per day and the pendulum had to be shortened. Dominique Cassini in Paris, himself an experienced observer, commented critically on the difficulties of keeping correct time on expeditions, and required extra care to be taken to overcome them. Nevertheless, observations by others confirmed Richer's experience. This supported Newton's deductions that the

centrifugal acceleration, due to the Earth's rotation, and an equatorial bulge diminish the intensity of gravity at the low latitudes.

The rotation and flattening of Jupiter were discovered by G. Cassini. Jupiter is more than ten times larger than the Earth and rotates more than twice faster. Consequently, Jupiter's flattening is large enough to be recognisable in the telescope.

1.8 Arc Measurements in Peru, Lapland and Some Others

It became evident that the collision between geometric and dynamic findings concerning the shape of the Earth could not be resolved by arc measurements at mid latitudes. The Académie Royale des Sciences decided to resolve this problem once and for all in a magnificently simple manner by sponsoring two famous expeditions, which made history and brought fame to the Academy.

In 1735, the first expedition, led by Pierre Bouguer and Charles Marie de la Condamine, was sent to Peru (now Ecuador) to measure the length of a meridional degree close to the equator. In 1736, the other expedition, led by Pierre L. M. Maupertuis, was sent to Lapland to make a similar measurement near the Arctic Circle. The latter measurement was carried out in a region north of the Gulf of Bothnia in the Baltic Sea (in the valley of the River Torneälven, which forms the present border between Sweden and Finland; see Fig. 1.6).

Several well-known scientists participated in these expeditions, which enhanced the scientific and public interest in these measurements. We shall become acquainted with the name of Bouguer later in connection with the Bouguer gravity reduction and Bouguer gravity anomaly. Maupertuis formulated one of the principles in analytical mechanics, now referred to as Maupertuis' principle. Another famous participant in the Lapland expedition was Celsius, who joined the expedition in Uppsala.

The places of measurement in Peru and Lapland were far enough apart for a decisive distinction between a prolate (French) and an oblate (British) ellipsoid. If a 1° meridional arc in Lapland were shorter than a 1° arc near the equator, then the French would be right. However, the measurements of these expeditions showed that the meridional degree in Lapland was longer, so that the British were right. So the expeditions confirmed the flattening of the Earth at the pole, as Newton had forecast. Maupertuis was praised upon his return by Voltaire for having "flattened the Earth and the Cassinis".

Some curious events also increased the dramatic character of the measurements. For example, before the beginning of the measurements, two identical standards of length were made in France, namely the Peru toise and Lapland toise. It was intended to compare them again after finishing the measurements. However, the Laplandian standard was lost as the ship carrying it sank on the return voyage.

The first arc measurements in the Southern Hemisphere went through a similar history as the measurements in the Northern Hemisphere, mainly due to

the development of the prolate into the oblate ellipsoid. Let us give a few details concerning this development.

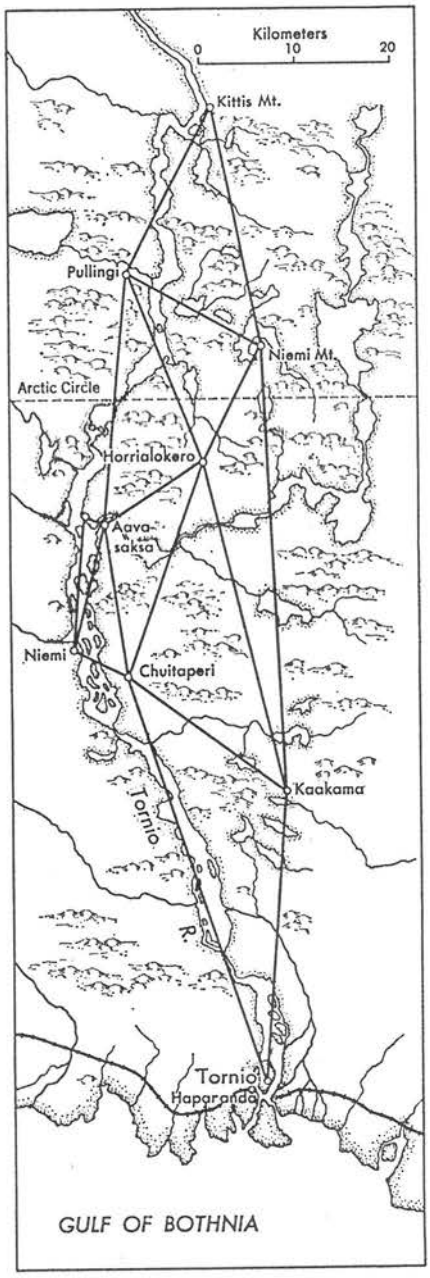

Figure 1.6: Arc measurement in Lapland 1736-1737 according to Outhier's map from the year 1736. (After Heiskanen and Vening Meinesz (1958)).

After completing the arc measurements in Lapland and Peru, the Académie Royale des Sciences wanted confirmation of the new concept also from the Southern Hemisphere. Consequently, in 1751, Nicolas de la Caille was sent to make arc measurements in South Africa. He measured a short arc from Cape Town north to Klipfontein, the length of which only slightly exceeded 1°. The result indicated a considerable difference in the hemispheres, the Southern Hemisphere being pointed towards the pole, i.e. having the shape of a prolate ellipsoid. Only much later, in 1837, Thomas Maclear revealed errors and large deflections of the vertical in the previous measurements. Maclear's measurement along a longer meridional arc confirmed the flattening of both hemispheres at the pole.

Even the famous arc measurements in Lapland were not free of serious errors, which could have had serious consequences (Tab. 1.2). The first value of the meridional flattening, computed by Maupertuis, was 1/216.8; now we know that it differs substantially from the correct value, which is 1/298.25. Maupertuis' error came to light in 1804 when the Swedish Geodetic Institute rechecked his result. It was found that Maupertuis' meridional degree was too large by some 440 m, due to an accumulation of errors which happened to be all in the same direction. On the one hand, this fortuitous error helped to settle the dispute about the figure of the Earth at the time. The thing was that the large flattening, found by Maupertuis, eliminated any doubts about the validity of Newton's theory. On the other hand, if the error had been in the other direction, the questions of the Earth's shape and the verification of Newton's theory might have remained unanswered and undecided.

At this moment, to be historically fair, we should reject the strong attacks against the Cassinis, mentioned above. Firstly, their conclusion about the egg-shaped Earth was based on measurements, not on hypotheses (and from the time of Galileo, the result of the experiment is considered to be the decisive criterion of validity of any physical theory). Secondly, also other arc measurements contained many inaccuracies and errors, as we have just seen. Many of these problems were caused by the irregular shape of the Earth, which was discovered later (see the next section).

The arc measurements in the 18th century, mainly the expeditions to Peru and Lapland, played a fundamental role in the development of geodesy and physics. Their significance is threefold:

1. They determined the figure of the Earth. Instead of one global parameter, the radius of the spherical Earth, two parameters were introduced, the semi-axes a, b of an ellipsoid of revolution, or the semi-major axis a and the flattening $\alpha = (a-b)/a$.
2. They confirmed Newton's theory of gravitation. Although Newton's theory explained Kepler's laws, independent proof of the theory was required. When the arc measurements in Peru and Lapland revealed the polar flattening of the Earth, Newton's theory was generally adopted. From this point of view, the arc measurements were even more important to physics than to geodesy. This important fact is not usually mentioned in

the present textbooks of physics. The prediction and discovery of the planet of Neptune is usually presented as a triumph and an excellent confirmation of the validity of the Newtonian theory of gravitation. However, the existence and position of this planet were predicted independently by Le Verrier and Adams in 1846, i.e. more than a century after the arc measurements by Bouguer and Maupertuis

3. Yet another physical aspect of the arc measurements is important, namely the metrological, because these measurements led to the first definition of the metre. On advice of the Académie Royale des Sciences, in 1791, the metre was defined as a $10\,000\,000^{\text{th}}$ part of the Earth's meridional quadrant. To determine it, J. P. J. Delambre and P. F. A. Méchain measured a long arc from Dunkerque to a point near Barcelona in 1792-1798. It was combined with the arc in Peru, which yielded a flattening of 1/334.29. The length of the meridional quadrant, computed in toises, was then set equal to 10 000 000 m by definition. Note that the length of the meridional quadrant L is related to equatorial radius a and flattening α as

$$L = a\frac{\pi}{2}\left(1-\frac{\alpha}{2}+\frac{\alpha^2}{16}-\ldots\right) . \tag{1.2}$$

The length of the equatorial radius in metres is then determined by this formula (Tab. 1.2).

Let us add several comments to the first and later definitions of the metre. The French original decision to establish such a natural unit, which would be rigid and recoverable for all times, was evidently inconsiderate and impractical. Even today we have problems in accurately determining the length of the quadrant. Luckily, a certain duplicate definition of the metre was introduced in 1799, when the resulting relationship between the metre and toise (Toise du Pérou) was fixed by law, and an accessible standard of the "legal" metre was deposited in the National Archives in Paris. The intended natural length unit was thus changed in essence to a conventional, the metre being defined as the distance between two thin lines on a length standard. For this second definition, the arc measurements actually became immaterial. This second definition of the metre was used for a long time, until the second half of the 20^{th} century. The historical connection with the size of the Earth was further removed in the next definition, in which the metre was defined in terms of the wavelength of a spectral line of the krypton-86 atom (introduced in connection with the International System of Units, SI). However, this definition was soon abandoned. At present, the length of the metre is derived from the velocity of light and the length of the second. The velocity of light in vacuum has now been adopted by definition to be $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$ precisely. This value of c, together with the definition of the second, then determine the length of the metre.

Table 1.2: The dimensions of the Earth ellipsoid: a is the equatorial radius, α is the flattening. The last column gives examples of the countries and regions where the corresponding ellipsoid was still in use in the 1970's or later.

Author	Year	a, m	$1/\alpha$	Used in
Bouguer, Maupertuis	1738	6 397 300	216.8	
Delambre, Méchain, Académie Royale	1799	6 375 730	334.29	
Everest	1830	6 377 276.345	300.801 7	India
Bessel	1841	6 377 397.155	299.152 8	China, Japan
Clarke	1866	6 378 206.4	294.978 7	North America
Modified Clarke	1880	6 378 249.145	293.465	Africa
Helmert (gravimetrically)	1901	-	298.3	
Helmert	1907	6 378 200	298.3	
Hayford, 1910, International Ellipsoid	1924	6 378 388	297.0	West Europe
Krasovskii	1940	6 378 245	298.3	Soviet Union
Buchar (Sputnik 2)	1958	-	297.9	
Reference Ellipsoid	1967	6 378 160	298.25	Australia South America
Geodetic Reference System 1980	1980	6 378 137	298.257	
I.A.G. proposal	1995	6 378 136.49	298.256 42	

Yet another fundamental contribution to geodesy and geophysics, which should be mentioned, was made in the 18th century. In 1743, Claude A. Clairaut derived a theorem which relates flattening α to the gravity variation and to the centrifugal acceleration at the equator, $\omega^2 a$:

$$\alpha + \frac{\gamma_p - \gamma_e}{\gamma_e} = \frac{5}{2}\frac{\omega^2 a}{\gamma_e}, \tag{1.3}$$

where γ_e and γ_p are the gravity accelerations at the equator and pole, ω is the angular velocity of the Earth's rotation. The derivation of this important theorem is described below in the chapter on the gravity field. The theorem enables the flattening to be determined from measurements of the gravity acceleration. Consequently, we already know two methods of determining the flattening of the Earth, namely the arc measurement and the application of Clairaut's theorem. The third fundamental method, the satellite method, will be mentioned further on in this chapter.

1.9 The Geoid as the Figure of the Earth

The famous French arc measurements were followed by similar measurements in other regions, and various numerical results were obtained. Inconsistencies in the observational data were often found when the two final unknowns, the equatorial radius and flattening, were to be determined from more than two equations. These puzzling problems attracted the attention of many famous contemporaneous mathematicians, and led to the development of new methods in mathematics. In particular, adjustment theories were developed at the beginning of the 19th century. The so-called method of least squares, independently developed by A. M. Legendre and C. F. Gauss, then became a basic tool in the processing of geodetic measurements.

Careful measurements indicated that the discrepancy in lengths or positions must be real, because it could not be explained by an accumulation of even extreme observational errors. Gauss recognized that the large residuals were caused by the irregular shape of the Earth, and proposed a new approach to these problems. In 1828 he wrote (Fischer, 1975): "The arc measurement in Hannover adds new confirmation to the now unquestionable truth that the surface of the Earth does not have a quite regular shape.... What we call the surface of the Earth in a mathematical sense, is nothing else but the surface, which everywhere intersects the direction of gravity at right angles, and of which the surface of the oceans is a part.

This does not prevent us, however, from considering the Earth as a whole to be a spheroid of revolution, from which the real (mathematical) surface deviates everywhere in larger or smaller, shorter or longer undulations... the ideal spheroid of revolution would be the one where the computed directions of the normals best agreed with the astronomic observations [today we would say 'where the deflections of the vertical are minimized']".

F. W. Bessel supported Gauss' ideas and elaborated them further. He analysed in detail the difference between the physical (actual) and the mathematical surface of the Earth. As the mathematical surface of the Earth he proposed to adopt an equipotential surface of the gravity field. In 1837 he wrote: "... irregularities of the mass distribution in the Earth's interior produce irregularities of the mathematical surface. All attractions together, combined with the centrifugal force produce that surface, to which the geodetic work refers ... one must still decide which of these surfaces should be the mathematical surface of the Earth. The choice would ... actually be arbitrary if the Earth were only a rigid body without an ocean. Since this, however, exists, it is appropriate to adopt that one as the surface of the Earth, of which the ocean surface is a part. Imagine the Earth covered by a net of channels connected with the ocean and filled by it, then the surface of the calm water in them would coincide with the mathematical surface of the Earth".

In these words, Bessel very clearly described the surface which we now call the geoid, and which we consider to be the mathematical surface of the Earth. The term "geoid" was introduced later on, in 1872, by J. B. Listing. At present,

the geoid is defined as that equipotential surface of the Earth's gravity field (attraction and rotation) which coincides best with the mean sea level.

In the conception of Gauss, the role of the mathematical surface of the Earth was transferred from the ellipsoid to the geoid, but the ellipsoid (or a spheroid) was retained as an approximation to the geoid. Gauss proposed to find a best-fitting world spheroid by minimizing the deflections of the vertical. Many authors attempted to derive the Earth ellipsoids in this way. Some results are given in Tab. 1.2. For more details we refer the reader to the lists in Heiskanen and Vening Meinesz (1958), Fischer (1975), and Grushinskii (1965, 1983).

The problem remained how to determine the distances between the geoid and ellipsoid. Astro-geodetic methods could solve this problem, but only on the continents. G. G. Stokes' theorem (1849) made it possible to solve this problem also on the oceans. It determined the distances between the geoid and ellipsoid from gravity observations. The deviations of the geoid from the ellipsoid are accompanied by deflections of the vertical from the normal to the ellipsoid. The formulae for computing these deflections from the same gravity observations were derived by F. A. Vening Meinesz (1928). These approaches accented the role of gravity measurements in solving the geodetic problems, and opened a new branch of science - physical geodesy.

A fundamental question, which has been repeatedly asked, is how accurately the geoid can be determined. The opinions of this problem varied. Bessel was rather sceptical when he wrote: "The extent of these undulations will not be known, unless made the specific purpose of a measurement ...". F. R. Helmert (1884) was more optimistic and emphasized that "we gained some knowledge of the figure of the Earth in general although its surface is not everywhere accessible". He ascribed this to "a very simple rule" of formation of the shape of the Earth under the influence of the gravitation and of the centrifugal force. In other words, theoretical analyses, based on the potential theory, should help us to overcome some technical problems. Despite the great progress in these investigations, some limitations of the classical approaches became more and more evident. M. S. Molodenskii (1945) came to the conclusion that it is impossible to determine the geoid only from measurements on the physical surface of the Earth. Later on, a new method was developed where the geoid, as a basic geodetic surface, was replaced by the so-called quasigeoid (Molodenskii et al., 1960). Discussions about these problems continue up to the present.

After the launching of the first man-made satellites of the Earth in 1957, a new epoch began in geodesy, geophysics and other branches of science and technology. Most of the present global data on the figure of the Earth and on its gravitational field were obtained by satellite techniques.

We shall not give further details of the development of physical geodesy, gravimetry, and satellite methods, since separate chapters are devoted to these problems below. Only a brief review of the main events is given in the next section.

1.10 A Review of Important Discoveries and Events

We shall only list some of the important events in the development of geodesy, gravimetry and study of the Earth's motions. For further details we refer the reader to Heiskanen and Vening Meinesz (1958), Pick et al. (1973), Fischer (1975), Grushinskii (1965, 1983), and Burša and Pec (1993). The events, which we have selected from the rich history of these investigations, are as follows:

6th century B.C. - Pythagoras - Concept of a spherical Earth.

3rd century B.C. - Eratosthenes - First successful measurement of the size of the Earth.

2nd century B.C. - Hipparchos - Discovery of precession.

1589 - Tycho Brahe - First application of triangulation to a measurement of distances.

1615 - Snellius - Introduction of triangulation into geodesy.

1669 - Picard - Arc measurement in northern France.

1672 - Richer - Change of the rate of pendulum clocks between Paris and Guyana.

1673 - Huygens - Theory of the physical pendulum; application to measurement of the free-fall acceleration.

1683-1716 - D. Cassini, J. Cassini - Arc measurement across the whole of France; non-sphericity of the Earth.

1687 - Newton - "Philosophiae naturalis principia mathematica" published; axiomatization of mechanics; law of universal gravitation; prediction of polar flattening of the Earth; explanation of precession and tides.

1735 - - French arc measurement expedition to Peru.

1736 - - French arc measurement expedition to Lapland.

1743 - Clairaut - Clairaut's theorem for determining the Earth's flattening from gravity observations.

1749 - Bouguer - Gravity in the Andes does not correspond to their mass.

1749 - D'Alembert - Analytical theory explaining precession and nutation.

1791 - - Definition of the metre as a 10 000 000th part of the meridional quadrant.

1792-1798 - Delambre, Méchain - First sufficiently accurate determination of the dimensions of the Earth.

1798 - Cavendish - Laboratory measurement of the gravitational constant.

1818 - Kater - Construction of the reversible pendulum; gravity measurements with this instrument.

1828 - Gauss - The irregular shape of the Earth was proved.

1830 - Everest - Everest's ellipsoid.

1837 - Bessel - The equipotential surface of the gravity field (now called the geoid) was proposed as the mathematical surface of the Earth.

1841 - Bessel - Bessel's ellipsoid.

1849 - Stokes - Stokes' theorem for computing the undulations of the geoid on the basis of gravity observations.

1851 - Foucault - Foucault's pendulum; demonstration of the Earth's rotation with the aid of a long pendulum at the Pantheon in Paris.
1854 - Pratt - Attraction of the Himalayas is partly compensated by a mass deficiency at depth; Pratt's hypothesis of isostasy.
1855 - Airy - Airy's hypothesis of isostasy.
1866 - Clarke - Clarke's ellipsoid.
1869 - Zöllner - Construction of the tiltmeter (Zöllner horizontal pendulum); measurement of tidal tilts of the vertical.
1872 - Listing - Introduction of the term "geoid".
1887 - Sterneck - Construction of a four-pendulum instrument for relative gravity measurements.
1889 - Dutton - Introduction of the term "isostasy".
1891 - Newcomb - Discrepancy between the Euler and Chandler periods explained by the elasticity of the Earth and motion of the waters of oceans and seas.
1895 - - International Latitude Service (ILS) established; regular observations began in 1899.
1898-1904 - Kühnen, Furtwängler - Absolute gravity measurements in Potsdam; basis of the Potsdam gravity system.
1901 - Helmert - Helmert's formula for normal gravity; gravimetric determination of the Earth's flattening yielded a value of 1/298.3, which is surprisingly close to the present value.
1910 - Eötvös - Construction of the gravitational torsion balance.
1911 - - Bureau International de l'Heure (BIH) founded in Paris.
1919 - - International Union of Geodesy and Geophysics (IUGG) established.
1924 - - Hayford's ellipsoid (1910) adopted as the International Ellipsoid (at the IUGG Assembly in Paris). H Madrid
1928 - Vening Meinesz - Vening Meinesz formulae for determining the deflections of the vertical on the basis of gravity observations.
1929 - Vening Meinesz - Pendulum gravity measurements at sea.
1930 - - Cassinis' formula for normal gravity adopted as the International Formula (IUGG Assembly in Madrid). ⊔ Stockholm
1932 - Tomaschek, Schaffernicht - Measurement of the tidal changes of the magnitude of gravity acceleration.
1940 - Krasovskii - Krasovskii's ellipsoid.
1945 - Molodenskii - Impossibility of determining the geoid from surface measurements.
1957 - - First man-made satellite of the Earth, Sputnik 1, launched on October 4.
1957 - - Sputnik 2 launched on November 3.
1958 - Buchar - First determination of the Earth's flattening from satellite measurements (from the motion of the orbital node of Sputnik 2). The value of the flattening, 1/297.9, differed significantly from the flattening of the International Ellipsoid (1/297.0).
1960 - Molodenskii et al.- Molodenskii's theory of the figure of the Earth.

1962 - - International Polar Motion Service (IPMS) established, replacing the ILS.

1967 - - "Geodetic Reference System 1967" (IUGG Assembly in Luzern, Switzerland); the first reference ellipsoid, the flattening of which was derived from satellite data.

1969 - - Laser corner reflectors installed on the Moon.

1976 - - LAGEOS (Laser Geodetic Satellite) launched.

1979 - - "Geodetic Reference System 1980" (IUGG Assembly in Canberra, Australia).

1988 - - International Earth Rotation Service (IERS) began operating on January 1 (its central bureau is in Paris), replacing the IPMS.

1995 - - A new set of fundamental parameters (proposed at the IUGG Assembly in Boulder, USA, but not yet adopted as a new reference system).

1.11 Present Values of Fundamental Constants

1.11.1 Explanation of some terms

Let us first explain certain differences between the terms "Earth ellipsoid", "reference ellipsoid" and "reference system", although some of them have already been used above.

The *Earth ellipsoid* is an ellipsoid approximating the geoid. The classical method of determining the parameters of such an ellipsoid is the arc measurement.

The *reference ellipsoid* is an Earth ellipsoid to which geodetic measurements, computations and mapping are referred. The parameters of the reference ellipsoid are often adopted on the basis of international agreements for a certain period of time, until they are improved.

The *flattening of the Earth* is the flattening of the corresponding Earth ellipsoid. In the classical geophysical literature, the flattening was usually denoted by the letter α, but recently it has often been replaced by the letter f. For an ellipsoid of revolution with the semi-axes a and b $(a > b)$, the flattening is defined as

$$\alpha = \frac{a-b}{a} .$$

A *reference system (or reference model)* is formed from the parameters of a reference ellipsoid by adding further global parameters. A standard reference model of the Earth can be defined by the following parameters:

- equatorial radius a,
- flattening α;
- mass of the Earth M;
- angular velocity of rotation ω.

Measurements by means of spacecraft and man-made satellites enabled little accurate values of the quantities M and α to be replaced by more accurate values of the geocentric gravitational constant GM and of a certain dimensionless parameter J_2 in the development of the gravitational potential, which is connected with the flattening; J_2 is the second-degree zonal geopotential (Stokes') parameter. The contemporary definition of a reference model is given by the following quantities: a, J_2, GM and ω. The flattening can be determined approximately from these quantities using the formula:

$$\alpha = \frac{3}{2} J_2 + \frac{\omega^2 a^3}{2GM} .$$

We shall derive this formula below in the chapter on the gravity field.

1.11.2 Geodetic Reference System 1980 and its modifications

The Geodetic Reference System 1980 was adopted at the end of the year 1979 by the IUGG General Assembly in Canberra, Australia. It is formed by the following constants:

$$\begin{cases} a = (6\,378\,137 \pm 2)\,\mathrm{m}; \\ \mathrm{J}_2 = (1\,082\,630 \pm 5) \times 10^{-9}; \\ GM = (398\,600.47 \pm 0.05) \times 10^9\ \mathrm{m^3 s^{-2}}; \\ \omega = 7\,292\,115 \times 10^{-11}\ \mathrm{rad\ s^{-1}}. \end{cases}$$

For the flattening we then get

$$\alpha = \frac{1}{(298.257 \pm 0.001)} .$$

The refined values of the above-mentioned constants (discussed internationally at the IUGG General Assembly in Boulder, USA, in 1995, but not yet adopted as a new reference system) are as follows:

$$\begin{cases} a = (6\,378\,136.49 \pm 0.10)\,\mathrm{m}; \\ \mathrm{J}_2 = (1\,082\,626.7 \pm 0.1) \times 10^{-9}; \\ GM = (398\,600\,441.8 \pm 0.8) \times 10^6\ \mathrm{m^3 s^{-2}}; \\ \omega = 7\,292\,115 \times 10^{-11}\ \mathrm{rad\ s^{-1}}. \end{cases}$$

For the flattening we then get

$$\alpha = \frac{1}{(298.256\,42 \pm 0.000\,01)} .$$

As a better approximation of the figure of the Earth it is possible to consider, instead of the ellipsoid of revolution, a tri-axial ellipsoid. Its flattening α_1 in the equatorial plane is small, $\alpha_1 = 1/(91\,026 \pm 10)$; the geographical longitude of the longer equatorial semi-axis is $\Lambda_a = (14.929° \pm 0.001°)$ W.

In addition to a, J_2, GM and ω, some other constants have been included in the geodetic reference systems. The principal constants are given in Tab. 1.3. Further details and references can be found in Moritz (1979) and Burša et al. (1995). Note that we should speak of "fundamental parameters" rather than of "fundamental constants", since small time variations of some parameters are involved (e.g., secular decreases of ω and J_2).

1.11.3 Methods of determining the fundamental parameters (a brief review)

The methods of determining some of the fundamental constants will be described in detail in the next chapters. Nevertheless, as a supplement to Tab. 1.3, let us also give a brief review of these methods here. Let us consider the individual parameters in the order as they appear in Tab. 1.3:

c : At present the velocity of light is postulated as an accurate constant, from which the length of the metre can be determined as a derived quantity (see Section 1.8).

G : The gravitational constant is usually determined with the aid of a torsion balance (Cavendish, 1798). Other methods are described in Sagitov (1969) and Stegena and Sagitov (1979).

GM :

1. *Approximately* this parameter can be determined from the mean value of gravity g and from the Earth's radius r (a spherically symmetric Earth):

$$g = \frac{GM}{r^2} .$$

Putting $g = 9.81\,\mathrm{m\,s^{-2}}$ and $r = 6370 \times 10^3\,\mathrm{m}$, we get $GM = 3.981 \times 10^{14}\,\mathrm{m^3 s^{-2}}$. (Here we have neglected the centrifugal acceleration.)

2. *More accurately* by means of *Gauss' law*:

$$\iint \mathbf{g} \cdot \mathbf{dS} = -4\pi GM .$$

3. Present methods are based on measurements of the motions of spacecraft and geodetic satellites.

ω : This parameter is determined astronomically.

J_2 : This is the second-degree zonal (Stokes') parameter. The gravitational potential can be expressed as

$$U = -\frac{GM}{r}\left[1 - \left(\frac{a}{r}\right)^2 J_2 P_2(\cos\theta) + \ldots\right],$$

where $P_2(\cos\theta) = \frac{3}{2}\cos^2\theta - \frac{1}{2}$, $\theta = 90° - \varphi$ (θ = polar distance, φ = geocentric latitude). Parameter J_2 can be determined from the motion of the orbital plane, or from the motion of the perigee of a man-made satellite (from perturbations of the orbits).

a : The classical determinations of the equatorial radius were based on arc measurements. At present, satellite data are also used.

γ_e : In the past this parameter was determined by averaging the surface gravity data to satisfy the formula for normal gravity:

$$\gamma = \gamma_e\left(1 + \beta\sin^2\varphi\right),$$

where φ is the latitude, γ_e and β are the constants to be determined. At present this parameter is derived from GM, ω, J_2, a.

α : In the past: arc measurements.
Later: Clairaut's theorem, i.e.

$$\alpha + \beta = \frac{5}{2}q, \quad q = \frac{\omega^2 a^3}{GM},$$

where β is the coefficient in the formula for normal gravity.
Now: satellite measurements (as a quantity derived from GM, ω, J_2, a; see Subsection 1.11.1)

α_1, Λ_a : From an approximation of the geoid by a tri-axial ellipsoid.

Table 1.3: The numerical values of the fundamental parameters in the Geodetic Reference System 1980 and their recent improvements.

Parameter	Symbol	Geodetic Reference System 1980	Refined value 1995
Velocity of light in vacuum	c	$(299\,792\,458 \pm 1.2)\,\mathrm{m\,s^{-1}}$	$299\,792\,458\,\mathrm{m\,s^{-1}}$ (accurate)
Newtonian gravitational constant	G	$(6\,672 \pm 4.1) \times 10^{-14}\ \mathrm{m^3 s^{-2} kg^{-1}}$	$(6\,672.59 \pm 0.30) \times 10^{-14}\ \mathrm{m^3 s^{-2} kg^{-1}}$
Geocentric gravitational constant (atmosphere included)	GM	$(398\,600.47 \pm 0.05) \times 10^{9}\ \mathrm{m^3 s^{-2}}$	$(398\,600\,441.8 \pm 0.8) \times 10^{6}\ \mathrm{m^3 s^{-2}}$
Mean angular velocity of the Earth's rotation	ω	$7\,292\,115 \times 10^{-11}\ \mathrm{rad\,s^{-1}}$	$7\,292\,115 \times 10^{-11}\ \mathrm{rad\,s^{-1}}$
Second-degree zonal geopotential parameter (without tides)	J_2	$(108\,263.0 \pm 0.5) \times 10^{-8}$	$(1\,082\,626.7 \pm 0.1) \times 10^{-9}$
Equatorial radius of the Reference Ellipsoid	a	$(6\,378\,137 \pm 2)\,\mathrm{m}$	$(6\,378\,136.49 \pm 0.10)\,\mathrm{m}$ IERS standard is 6 378 136 m
Mean equatorial gravity	γ_e	$(978\,033 \pm 1) \times 10^{-5}\ \mathrm{m\,s^{-2}}$	$(978\,032.78 \pm 0.2) \times 10^{-5}\ \mathrm{m\,s^{-2}}$
Polar flattening	α	$1/(298.257 \pm 0.001)$	$1/(298.256\,42 \pm 0.000\,01)$
Equatorial flattening	α_1	1/90 000	$1/(91\,026 \pm 10)$
Longitude of major axis of equatorial ellipse	Λ_a	15° W	(14.9291°±0.0010°) W

Chapter 2

Motions of the Earth - Part I

The Earth is a part of the Solar System which is a part of the Galaxy. Together with these systems, the Earth performs a series of motions. We shall, therefore, start the description of the motions of the Earth with the motions with respect to remote astronomical surroundings. Then we shall proceed to motions of smaller scales.

2.1 Motions with Respect to Background Radiation

The background radiation is the microwave radiation of the Universe at frequencies of 10^8 to 10^{10} Hz. The radiation has all the characteristics of the radiation of a black body at a temperature slightly below 3 K $(2.735 \pm 0.060\,\mathrm{K})$. It is considered to be a remnant and one of the main proofs of the existence of the Big Bang $\left((9-16)\times 10^9 \text{ years ago}\right)$. The radiation was predicted by G. Gamow in 1948, and discovered by A. A. Penzias and R. Wilson in 1965. Penzias and Wilson were awarded the Nobel Prize for this discovery in 1978.

Observations from airplanes (to suppress the radiation coming from the atmosphere) revealed a small anisotropy of this microwave radiation. If we consider the reference frame in which the background radiation is isotropic, then the Solar System, the Earth and the instrument in the airplane move with respect to this reference frame at a velocity of about $400\,\mathrm{km\,s^{-1}}$. (At this accuracy of measurements we cannot distinguish the motion of the Earth round the Sun). If the known motion of the Solar System in the Galaxy is taken into account, we can determine the motion of the central region of the Galaxy with respect to this selected reference frame. We obtain a velocity of about $600\,\mathrm{km\,s^{-1}}$ toward the constellation of Sextant.

The selected reference frame, in which the background radiation is isotropic, resembles, in some features, Newton's absolute reference frame, the existence of which was rejected by the theory of relativity. Does the existence of this selected reference frame mean that we should revise our relativistic theories? We do not know yet (Kalvoda, 1981).

2.2 Other Motions of the Galaxy

The Galaxy moves with respect to the local group of galaxies at a velocity of about $100\,\mathrm{km\,s^{-1}}$. This group of galaxies moves toward the cluster of galaxies in the constellation of Virgin at a velocity of about $1000\,\mathrm{km\,s^{-1}}$.

2.3 Rotation of the Galaxy

The central parts of the Galaxy rotate faster than the parts on its periphery. The distance of the Solar System from the centre of the Galaxy is about 7 kpc. (Note that 1 light year $= 9.46 \times 10^{12}$ km $= 0.307$ pc). Thus 7 kpc is about 23 000 light years. The period of revolution of the Solar System is about 200 million years (190 million years is a more accurate value), and the velocity of revolution is about $230\ \mathrm{km\,s^{-1}}$.

Two interesting facts should be mentioned, because they may indicate broader consequences:

1. There is an apparent coincidence between the period of the revolution of the Solar System (200 million years) and the period of some geological processes on the Earth. A possible explanation of this coincidence is still lacking.
2. The position of the Solar System is close to the so-called co-rotation orbit (the orbit where the velocity of revolving stars is the same as the velocity of density waves which form the spiral arms of the Galaxy). This position might play an important role in the origin of the Solar System.

2.4 Motion of the Solar System

This motion is understood to be the motion with respect to the nearest vicinity in the Galaxy, i.e. with respect to the nearest stars. The Solar System moves toward the constellation of Hercules at a velocity of about $20\ \mathrm{km\,s^{-1}}$ (determined spectroscopically from the Doppler effect). This motion can be considered approximately as a straight-line motion.

2.5 Revolution of the Earth round the Sun

The orbit of the Earth round the Sun is called the ecliptic (Fig. 2.1). The well-known motion along this orbit is characterized by the following parameters:

Mean distance from the Sun: 149.598×10^6 km $(= 1\ \mathrm{AU})$.

Period of revolution: $365^{\mathrm{d}}5^{\mathrm{h}}48^{\mathrm{m}}46^{\mathrm{s}}\left(\sim 365\,\tfrac{1}{4}\ \mathrm{days}\right)$.

Mean orbital velocity: $29.8\ \mathrm{km\,s^{-1}}$.

Eccentricity $e = 0.016\ 74$ (nearly circular orbit).

Inclination of the Earth's axis: $23°26'21''\left(\sim 23\,\tfrac{1}{2}°\right)$; this causes the rotation of the seasons.

Time of the passage through the perihelion (the closest point of the ecliptic to the Sun): on or about January 2.

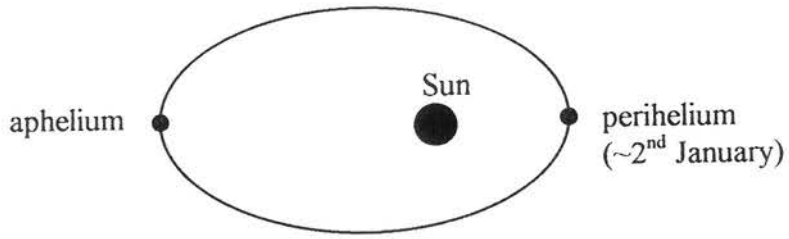

Figure 2.1: Ecliptic.

2.6 Rotation of the Earth

The second is defined so that the mean solar day has 86 400 seconds. However, one rotation (about 360°) takes 86 164.09 s (sidereal day). Thus the solar day has 86 400 s, whereas the sidereal day has 86 164.09 s.

The difference between the length of the solar and sidereal days is explained in Fig. 2.2. Consider a point P on the Earth's surface at noon on one day. After completing a rotation of 360° (sidereal day), it will not be noon at point P, but a certain period of time will be required to complete the solar day. Consequently, one year has 365 solar days, but 366 sidereal days.

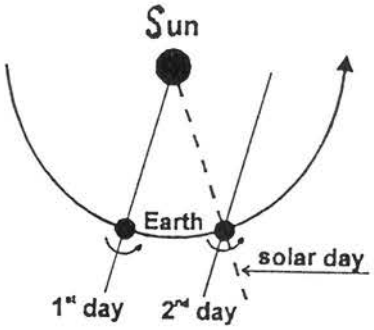

Figure 2.2: Explanation of the difference between solar and sidereal days. The orbital and rotational motions of the Earth are indicated by arrows (the inclination of the Earth's axis is ignored).

Remember that the *angular velocity* (angular frequency) of the Earth's rotation is

$$\omega = \frac{2\pi}{86\,164.09} = 7.292\,115 \times 10^{-5}\ \mathrm{s}^{-1}\ .$$

The velocity of a point on the equator due to the Earth's rotation is $465\ \mathrm{m\,s}^{-1}$, and the centrifugal acceleration on the equator is $3.39\ \mathrm{cm\,s}^{-1}$.

2.6.1 Consequences and proofs of the Earth's rotation

Astronomical observations give more or less indirect proof of the Earth's rotation. But the rotation can be demonstrated directly by experiments carried out on the surface of the Earth. The main experiments and phenomena are as follows:

a) The Foucault pendulum (shifting of the plane of the pendulum). This experiment was first performed publicly by Foucault in 1851 (Pantheon in Paris, a 28 kg sphere on a wire suspension 67 m long). The angle of the shift $\alpha = (\omega \sin\phi) \cdot t$, where ω is the angular velocity of the Earth's rotation, ϕ latitude, t time. Thus, the shift depends on latitude; it is maximum at the poles and zero at the equator (Fig. 2.3).

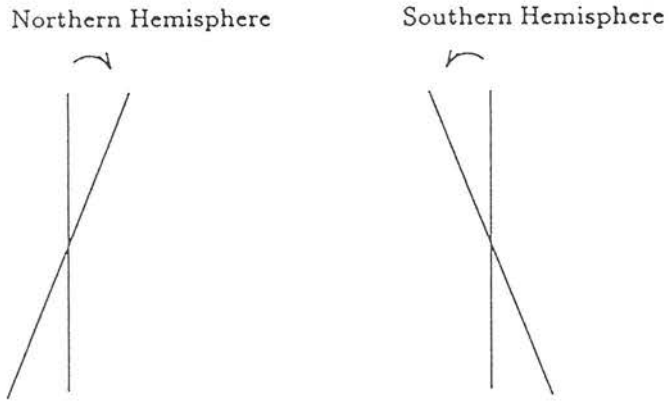

Figure 2.3: Change of the plane of motion of the Foucault pendulum.

b) Eastward deviation in free fall.
c) Deflections of the trade winds, sea currents, and others (Fig. 2.4).

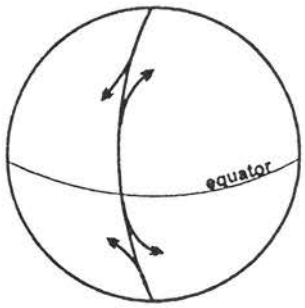

Figure 2.4: Deflections of motions on the rotating Earth due to the Coriolis force.

d) Decrease in weight when moving from west to east. This effect can be demonstrated with the aid of a rotating balance.
e) Gyroscopes etc.

Other motions of the Earth are connected with the time variations of its rotation. Before describing them in Chapter 4, we shall study the effects of the Earth's rotation on mechanical processes on the Earth in the next chapter.

Chapter 3

Mechanics in Non-Inertial Reference Frames

Physical measurements and observations are usually carried out in the reference frames which are fixed with the Earth. Due to its rotation, such reference frames are non-inertial. This affects the description of many processes studied in physics, geophysics, meteorology, astronomy and other natural sciences. As theses problems are usually mentioned in standard textbooks of physics only marginally, we shall devote more attention to them in this chapter. From a common point of view we shall derive the equation of motion of a particle in a non-inertial reference frame, Euler's dynamic equations for a rigid body, and Liouville's equations for a partly deformable body. Applications of these equations in geophysics and meteorology will be discussed in this and in the following chapters.

In preparing this chapter we have drawn on the textbooks by Kittel et al. (1962), Matveev (1986), Trkal (1956) and some others.

3.1 Time Variation of an Arbitrary Vector in Different Coordinate Systems. Resal's Theorem

Consider two coordinate systems, one of which will be considered fixed. Assume that the second system rotates, without translation, with respect to the fixed system at angular velocity ω.

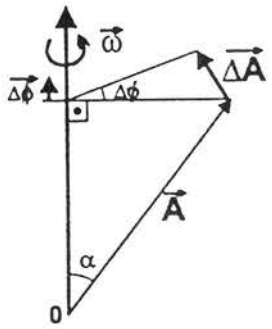

Figure 3.1: Time variation of a vector **A** which is constant in the rotating system.

Firstly, consider an arbitrary vector **A** which is constant in the rotating system (Fig. 3.1); for example, the radius-vector of any point of the rotating body. We wish to determine the time variation of this vector as seen from the fixed system. In a short time interval Δt, vector **A**, observed from the fixed system, changes by vector $\Delta\mathbf{A}$, the magnitude of which is

$$|\Delta\mathbf{A}| = \rho\,\Delta\phi = |\mathbf{A}|\sin\alpha\;\Delta\phi = |\Delta\boldsymbol{\phi}\times\mathbf{A}|\ ,$$

where ρ is the distance of the endpoint of **A** from the rotation axis, $\Delta\phi$ is the corresponding angle of rotation. Taking the directions of the vectors $\Delta\mathbf{A}$ and $\Delta\boldsymbol{\phi}$ into account, it can be seen that

$$\Delta\mathbf{A} = \Delta\boldsymbol{\phi}\times\mathbf{A}\ .$$

Dividing this by Δt yields

$$\frac{\Delta\mathbf{A}}{\Delta t} = \frac{\Delta\boldsymbol{\phi}}{\Delta t}\times\mathbf{A}\ .$$

As Δt approaches zero, and considering $\boldsymbol{\omega} = \Delta\boldsymbol{\phi}/\mathrm{d}t$, we arrive at

$$\boxed{\frac{\Delta\mathbf{A}}{\Delta t} = \boldsymbol{\omega}\times\mathbf{A}}\ . \tag{3.1}$$

This formula yields the time derivative, as seen from the fixed system, of a vector which is constant in the rotating system.

Secondly, let us generalise formula (3.1) also with respect to vectors which are not constant in the rotating system. Thus, assume that vector **A** is variable in the rotating system. Denote **i, j, k** the unit vectors of a Cartesian system which is fixed in the rotating system, i.e. these vectors are constant in the rotating system, and have the following components there:

$$\mathbf{i} = (1,0,0)\,;\quad \mathbf{j} = (0,1,0)\,;\quad \mathbf{k} = (0,0,1)\ . \tag{3.2}$$

Decompose vector **A** in the rotating system as follows:

$$\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k}\ , \tag{3.3}$$

where A_x, A_y, A_z are scalars. The time derivative of **A** with respect to the rotating system (time variation as seen in the rotating system) is

$$\left(\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right)_{rot} = \frac{\mathrm{d}A_x}{\mathrm{d}t}\mathbf{i} + \frac{\mathrm{d}A_y}{\mathrm{d}t}\mathbf{j} + \frac{\mathrm{d}A_z}{\mathrm{d}t}\mathbf{k}\ ; \tag{3.4}$$

the derivatives of the unit vectors in (3.3) are equal to zero as these vectors are constant. Now let us calculate the time derivative of **A** with respect to the fixed system. As vectors (3.2) are constant in the rotating system, we can use (3.1) to obtain:

$$\left(\frac{\mathrm{d}\,\mathbf{i}}{\mathrm{d}\,t}\right)_{fix} = \omega \times \mathbf{i}\ , \quad \left(\frac{\mathrm{d}\,\mathbf{j}}{\mathrm{d}\,t}\right)_{fix} = \omega \times \mathbf{j}\ , \quad \left(\frac{\mathrm{d}\,\mathbf{k}}{\mathrm{d}\,t}\right)_{fix} = \omega \times \mathbf{k}\ . \tag{3.5}$$

Differentiating (3.3) again, we now arrive at

$$\left(\frac{\mathrm{d}\,\mathbf{A}}{\mathrm{d}\,t}\right)_{fix} = \frac{\mathrm{d}\,A_x}{\mathrm{d}\,t}\mathbf{i} + \frac{\mathrm{d}\,A_y}{\mathrm{d}\,t}\mathbf{j} + \frac{\mathrm{d}\,A_z}{\mathrm{d}\,t}\mathbf{k} + A_x(\omega \times \mathbf{i}) + A_y(\omega \times \mathbf{j}) + A_z(\omega \times \mathbf{k}) =$$

$$= \left(\frac{\mathrm{d}\,\mathbf{A}}{\mathrm{d}\,t}\right)_{rot} + \omega \times \left(A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k}\right)\ ,$$

where we have used (3.5) and (3.4). Consequently, we arrive at the final formula:

$$\boxed{\left(\frac{\mathrm{d}\,\mathbf{A}}{\mathrm{d}\,t}\right)_{fix} = \left(\frac{\mathrm{d}\,\mathbf{A}}{\mathrm{d}\,t}\right)_{rot} + \omega \times \mathbf{A}}\ . \tag{3.6}$$

This is the basic formula relating the time variations of a vector in two coordinate systems. It is also called *Resal's theorem* (Resal, 1884; Burša and Pec, 1993). We shall use it many times in the following text.

To suppress the subscripts in formula (3.6), we shall also put

$$\left(\frac{\mathrm{d}\,\mathbf{A}}{\mathrm{d}\,t}\right)_{fix} = \dot{\mathbf{A}}$$

and omit the subscript "rot". Equation (3.6) then takes the form

$$\boxed{\dot{\mathbf{A}} = \frac{\mathrm{d}\,\mathbf{A}}{\mathrm{d}\,t} + \omega \times \mathbf{A}}\ . \tag{3.7}$$

As a special case of this formula, by substituting ω for $\mathbf{A}$, we have

$$\dot{\omega} = \frac{\mathrm{d}\,\omega}{\mathrm{d}\,t}\ . \tag{3.8}$$

This indicates that the time variation of vector ω is the same in both systems.

It should be mentioned that the fixed system in this Section 3.1 need not be an inertial system; the formulae here are quite general. However, in considering the equations of motion in the next sections, we shall usually assume that the fixed system is inertial.

3.2 Equation of Motion of a Particle in a Non-Inertial System

Again consider two coordinate system and assume that:

1. The fixed system is inertial (for Newton's second law to hold true);
2. The origins of the systems are generally different.

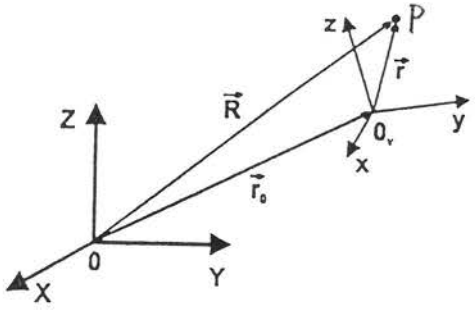

Figure 3.2: Position of point P in the two systems: **R** and **r** are the radius-vectors in the inertial and non-inertial systems, respectively.

Denote the Cartesian axes in the fixed system X, Y, Z, and in the moving system x, y, z (Figure 3.2). Denote $\mathbf{r}_0$ the radius-vector of the origin of the moving system with respect to the fixed system. The position of an arbitrary point P can be described by two radius-vectors, by **R** in the fixed system, and by **r** in the moving system, where

$$\mathbf{R} = \mathbf{r}_0 + \mathbf{r} \ . \tag{3.9}$$

We shall need the second derivative of this vector. In the first derivative,

$$\dot{\mathbf{R}} = \dot{\mathbf{r}}_0 + \dot{\mathbf{r}} \ , \tag{3.10}$$

the terms on the right-hand side could be expressed by means of Resal's theorem (3.7) as follows:

$$\dot{\mathbf{r}}_0 = \frac{\mathrm{d}\,\mathbf{r}_0}{\mathrm{d}\,t} + \boldsymbol{\omega} \times \mathbf{r}_0 = \mathbf{v}_0 \ , \tag{3.11}$$

$$\dot{\mathbf{r}} = \frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,t} + \boldsymbol{\omega} \times \mathbf{r} \ . \tag{3.12}$$

However, we shall not usually apply Resal's theorem to $\dot{\mathbf{r}}_0$ (we shall not describe the time variation of $\mathbf{r}_0$ from the moving system, as this would usually

be impractical). We shall simply write $\dot{\mathbf{r}}_0 = \mathbf{v}_0$, where $\mathbf{v}_0$ is the absolute velocity of the origin of the non-inertial system. Consequently,

$$\boxed{\dot{\mathbf{R}} = \mathbf{v}_0 + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{r}} \,. \tag{3.13}$$

In calculating the second derivative of $\mathbf{R}$ we shall apply Resal's theorem to all terms of (3.13):

$$\ddot{\mathbf{R}} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\mathbf{v}_0 + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{r}\right] + \boldsymbol{\omega} \times \left[\mathbf{v}_0 + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{r}\right] =$$

$$= \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + 2\boldsymbol{\omega} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{v}_0}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{v}_0 + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \,.$$

Newton's second law in the inertial system,

$$m\ddot{\mathbf{R}} = \mathbf{F} \,, \tag{3.14}$$

then yields the equation of motion in the non-inertial system in the following form:

$$\boxed{m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \mathbf{F} - 2m\boldsymbol{\omega} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} - m\left[\frac{\mathrm{d}\mathbf{v}_0}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{v}_0 + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})\right]} \,. \tag{3.15}$$

The latter equation resembles the equation of motion in an inertial system (3.14), but in addition to the true force $\mathbf{F}$, there are two other terms, the so-called fictitious forces, on the right-hand side of the equation of motion:

$$\boxed{m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \mathbf{F} + \mathbf{F}_c + \mathbf{F}_d} \,. \tag{3.16}$$

The force

$$\mathbf{F}_c = -2m\boldsymbol{\omega} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = 2m\mathbf{v}_r \times \boldsymbol{\omega}$$

is called the Coriolis force, $\mathbf{v}_r = \dfrac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}$ being the velocity relative to the non-inertial system. Force $\mathbf{F}_d$, representing the last term in equation (3.15), is called the drifting force.

Note that the Coriolis force disappears if:

1) $\dfrac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = 0 \quad (\mathbf{v}_r = 0)$, or

2) $\dfrac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}$ is parallel to ω $\left(\mathbf{v}_r \text{ is parallel to } \omega\right)$.

Consider a very important special case when both the origins of the coordinate systems coincide, $O = O_v$. Then $\mathbf{v}_0 = 0$ and the equation of motion (3.15) simplifies to

$$\boxed{m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \mathbf{F} - 2m\boldsymbol{\omega} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} - m\dot{\boldsymbol{\omega}} \times \mathbf{r} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})} \,. \qquad (3.17)$$

The last term but one in the latter equation is called the Euler force,

$$\mathbf{F}_e = -m\dot{\boldsymbol{\omega}} \times \mathbf{r} \,,$$

but this term disappears if the angular velocity of rotation ω is a constant vector. In considering mechanical phenomena on the rotating Earth, this term is usually neglected, as it is smaller than the other terms. Nevertheless, some changes of the angular velocity of the Earth's rotation will be discussed in the next chapter.

The last term in equation (3.17) represents the centrifugal force, which can easily be verified as follows. From the vector relation

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

it follows that

$$\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = \boldsymbol{\omega}(\boldsymbol{\omega} \cdot \mathbf{r}) - \mathbf{r}(\boldsymbol{\omega} \cdot \boldsymbol{\omega}) = \omega^2\left[\frac{\boldsymbol{\omega}}{\omega}\left(\frac{\boldsymbol{\omega}}{\omega} \cdot \mathbf{r}\right) - \mathbf{r}\right] = -\rho\omega^2 \,.$$

Since $\dfrac{\boldsymbol{\omega}}{\omega}$ is the unit vector at the rotation axis, term $\dfrac{\boldsymbol{\omega}}{\omega}\left(\dfrac{\boldsymbol{\omega}}{\omega} \cdot \mathbf{r}\right)$ is the projection of $\mathbf{r}$ onto this axis (Fig. 3.3). Consequently, ρ is the oriented distance from the rotation axis, and $-\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = +\omega^2\rho$ is the centrifugal acceleration.

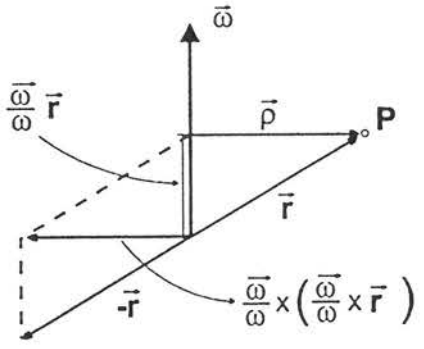

Figure 3.3: Explanation of the formula for the centrifugal acceleration.

Note that in considering a non-inertial reference system with its origin on the Earth's surface, we must use general equation (3.15), not (3.17). For example we have just seen that term $[-\omega \times (\omega \times \mathbf{r})]$ yields the whole centrifugal acceleration only if vector $\mathbf{r}$ is measured from a point on the rotation axis.

Let us go back to formula (3.13) and to its interpretation. This formula represents the time derivative of the sum (3.9) of two vectors. However, formula (3.13) can also be interpreted as a certain generalisation of Resal's theorem (3.7) for the case when also a translation of the moving system is considered. We shall return to this comment below in Section 3.5.

3.3 Equations of Motion of a System of Particles in an Inertial System

The radius-vector of the centre of mass of a system of particles is defined as (Fig. 3.4)

$$\mathbf{R} = \frac{\sum_{i=1}^{n} m_i \mathbf{r}_i}{\sum_{i=1}^{n} m_i} = \frac{\sum_{i=1}^{n} m_i \mathbf{r}_i}{M},$$

where m_i is the mass of the i-th particle, $\mathbf{r}_i$ its radius-vector, and M the total mass of the system.

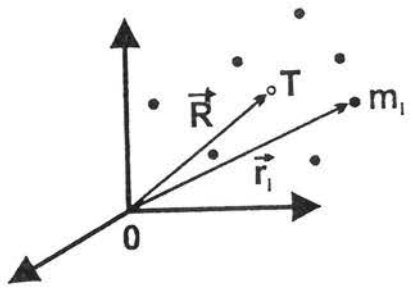

Figure 3.4: A system of particles, and their centre of mass, T.

The equation of motion of the i-th particle, following from Newton's second law, can be expressed as

$$\boxed{m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \mathbf{F}_i^{(e)} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbf{F}_{ij}}, \quad i = 1, 2, \ldots, n, \tag{3.18}$$

where $\mathbf{F}_i^{(e)}$ is the external force acting on the particle, and $\mathbf{F}_{ij}$ is the internal force due to the j-th particle. Equations (3.18) represent n vector equations which describe the system of n particles completely.

We shall now derive several special equations which follow from these equations.

3.3.1 Impulse-momentum theorem

The summation of all equations (3.18) yields

$$\sum_{i=1}^{n} m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \sum_{i=1}^{n} \mathbf{F}_i^{(e)} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbf{F}_{ij} \ , \tag{3.19}$$

which can be expressed as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(\sum_{i=1}^{n} m_i \mathbf{r}_i\right) = \sum_{i=1}^{n} \mathbf{F}_i^{(e)} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbf{F}_{ij} \ .$$

Assuming the validity of Newton's third law, $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$, the resultant of the internal forces vanishes, so that

$$\boxed{M \frac{\mathrm{d}^2 \mathbf{R}}{\mathrm{d}t^2} = \sum_{i=1}^{n} \mathbf{F}_i^{(e)}} \ . \tag{3.20}$$

Accordingly, the centre of mass moves as if the total mass of the system were concentrated there, and the resultant of the external forces were acting on it.

Let us introduce other notations:

- linear momentum $\mathbf{p}_i = m_i \mathbf{v}_i = m_i \dfrac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}$;
- the total linear momentum of the system,
  $\mathbf{P} = \sum_{i=1}^{n} \mathbf{p}_i = \sum_{i=1}^{n} m_i \mathbf{v}_i = \sum_{i=1}^{n} m_i \dfrac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}$;
- and the resultant of the external forces, $\mathbf{F}^{(e)} = \sum_{i=1}^{n} \mathbf{F}_i^{(e)}$.

Equation (3.19) can then be modified to read

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{n} m_i \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}\right) = \sum_{i=1}^{n} \mathbf{F}_i^{(e)} + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbf{F}_{ij} \ ,$$

and (3.20) takes the form

$$\boxed{\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \mathbf{F}^{(e)}} \ . \tag{3.21}$$

In this way we have derived the following important theorem:

Impulse-momentum theorem: *The time variation of the total linear momentum of a system is equal to the resultant of the external forces acting on the system if the internal forces obey Newton's third law.*

In the special case of $\mathbf{F}^{(e)} = 0$, vector $\mathbf{P}$ is constant, which yields the conservation of linear momentum.

3.3.2 Angular momentum theorem

Start again with (3.18) and multiply it from the left (vector multiplication) by radius-vector $\mathbf{r}_i$:

$$\mathbf{r}_i \times m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \mathbf{r}_i \times \mathbf{F}_i^{(e)} + \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbf{r}_i \times \mathbf{F}_{ij} \ . \tag{3.22}$$

Let us introduce the following notations:

- angular momentum $\mathbf{l}_i = \mathbf{r}_i \times \mathbf{p}_i = \mathbf{r}_i \times m_i \mathbf{v}_i$;
- total angular momentum of the system, $\mathbf{L} = \sum_{i=1}^{n} \mathbf{l}_i = \sum_{i=1}^{n} \mathbf{r}_i \times m_i \mathbf{v}_i$;
- total external torque $\mathbf{N}^{(e)} = \sum_{i=1}^{n} \mathbf{r}_i \times \mathbf{F}_i^{(e)}$ (turning momentum).

Assuming that the internal forces satisfy Newton's third law, $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$,

$$\mathbf{r}_i \times \mathbf{F}_{ij} + \mathbf{r}_j \times \mathbf{F}_{ji} = \left(\mathbf{r}_i - \mathbf{r}_j\right) \times \mathbf{F}_{ij} \ .$$

If $\mathbf{F}_{ij}$ is parallel to $\left(\mathbf{r}_i - \mathbf{r}_j\right)$, e.g., $\mathbf{F}_{ij}$ is a central force, the internal torques vanish in summing the equations. The summation of equations (3.22) then yields

$$\boxed{\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \mathbf{N}^{(e)}} \ . \tag{3.23}$$

We can formulate this resulting equation as follows:

Angular momentum theorem: *The time rate of change of the total angular momentum of a system is equal to the total external torque if the internal forces are central and obey Newton's third law.*

Let us add three remarks to these theorems:

1. The theorems were *derived*, under certain assumptions, from Newton's second law.
2. The impulse-momentum and angular momentum theorems represent two vector equations only. They are *not sufficient* to describe the motions of the system of n particles completely; other $n-2$ equations are needed.
3. The origin of the coordinate system, O, was arbitrary.

3.4 Equations of Motion of a Rigid Body in an Inertial System

We *postulate* that the motion of a rigid body is *completely* described by the impulse-momentum and angular momentum theorems (i.e. these theorems represent its equations of motions). This can be verified by experiments. Thus, we shall express the equations of motion of a rigid body in the form

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \mathbf{F} \quad \text{or} \quad M\frac{\mathrm{d}^2\mathbf{R}}{\mathrm{d}t^2} = \mathbf{F}\ , \tag{3.24a}$$

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \mathbf{N}\ , \tag{3.24b}$$

where superscript $^{(e)}$ has been omitted and $\mathbf{F} = \mathbf{F}^{(e)}$ is the resultant of the external forces and constraints.

Consider the special case when the body is fixed at a point. Then any motion is a rotation. Consider the fixed point to be the origin of a non-inertial system which is fixed in the rigid body. The radius-vector of the centre of mass, $\mathbf{R}$, is then constant in the non-inertial system, so that

$$\left(\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}\right)_{inert} = \boldsymbol{\omega} \times \mathbf{R}\ ,$$

and equation (3.24a) takes the form

$$M\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{\omega} \times \mathbf{R}) = \mathbf{F}\ ,$$

which is the equation of motion of the centre of mass (a first-order differential equation only).

3.5 Equations of Motion of a Rigid Body in a Non-Inertial System. Euler's Equations

Equations (3.24a) and (3.24b) using Resal's theorem (3.7) yield

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{P} = \mathbf{F} \ , \tag{3.25a}$$

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} + \boldsymbol{\omega} \times \mathbf{L} = \mathbf{N} \ . \tag{3.25b}$$

These equations are the general equations of motion of the rigid body in a non-inertial system, which do not require the rigid body to be fixed in the system. On the other hand, the origin of the non-inertial system must be fixed in an inertial system, since no translations are considered in Resal's theorem. (In other words, here we have used Resal's theorem (3.7) but not a more general equation of type (3.13)).

As a special case, let us assume that the rigid body is fixed at a point (see Section 3.4). In equation (3.24b) describing the rotation we shall put

$$L_i = J_{ik}\omega_k \ , \tag{3.26}$$

where L_i is the i-th component of the angular momentum and J_{ik} are the inertial coefficients. (Here we have taken equation (3.26) without any proof, assuming that the reader knows it from a course of physics. It will be derived, even for a more general case, in Section 3.6). Let us add another assumption that the axes, fixed in the body, coincide with the principal axes. Then

$$L_x = A\omega_x \ , \quad L_y = B\omega_y \ , \quad L_z = C\omega_x \ . \tag{3.27}$$

where A, B, C are the moments of inertia, respectively. By substituting these expressions into the x-component of equation (3.25b), we arrive at

$$\frac{\mathrm{d}L_x}{\mathrm{d}t} + \omega_y L_z - \omega_z L_y = N_x \ ,$$

or

$$A\frac{\mathrm{d}\omega_x}{\mathrm{d}t} - (B-C)\omega_y\omega_z = N_x \ .$$

If analogous equations are used to expressed the remaining two components, we finally arrive at *Euler's dynamic equations*:

$$A\frac{\mathrm{d}\omega_x}{\mathrm{d}t}-(B-C)\omega_y\omega_z = N_x \ ,$$

$$B\frac{\mathrm{d}\omega_y}{\mathrm{d}t}-(C-A)\omega_z\omega_x = N_y \ , \tag{3.28}$$

$$C\frac{\mathrm{d}\omega_z}{\mathrm{d}t}-(A-B)\omega_x\omega_y = N_z \ .$$

3.6 More General Equations of Motion in a Non-Inertial System. Liouville's Equations

Consider a body which is not absolutely rigid, for example, the solid Earth with its fluid parts, such as the oceans, atmosphere or the outer core. We wish to generalise Euler's equations for such a body.

Let us return to the definition of the total angular momentum of a system of particles (Section 3.3):

$$\mathbf{L} = \sum \mathbf{r}_i \times m_i \mathbf{v}_i = \sum \mathbf{r}_i \times m_i \dot{\mathbf{r}}_i \ .$$

Applying Resal's theorem (3.7) and putting $\mathbf{u}_i = \mathrm{d}\mathbf{r}_i/\mathrm{d}t$ yields

$$\mathbf{L} = \sum m_i \mathbf{r}_i \times \mathbf{u}_i + \sum m_i \mathbf{r}_i \times \left(\boldsymbol{\omega} \times \mathbf{r}_i\right) \ .$$

Rearrangement of the last term, using the identity

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \ ,$$

yields

$$\mathbf{L} = \sum m_i \mathbf{r}_i \times \mathbf{u}_i + \boldsymbol{\omega} \sum m_i \mathbf{r}_i^2 - \sum m_i \mathbf{r}_i \left(\mathbf{r}_i \cdot \boldsymbol{\omega}\right) \ .$$

For a body with a spatial distribution of mass density ρ, analogously

$$\mathbf{L} = \iiint_V \rho \, (\mathbf{r} \times \mathbf{u}) \, \mathrm{d}V + \boldsymbol{\omega} \iiint_V \rho \, \mathbf{r} \cdot \mathbf{r} \, \mathrm{d}V - \iiint_V \rho \, \mathbf{r} (\mathbf{r} \cdot \boldsymbol{\omega}) \, \mathrm{d}V \ .$$

The i-th component of the total angular momentum

$$L_i = \iiint_V \rho \in_{ijk} x_j u_k \, \mathrm{d}V + \omega_i \iiint_V \rho x_k x_k \, \mathrm{d}V - \iiint_V \rho x_i x_j \omega_j \, \mathrm{d}V \ ,$$

where we have used the Levi-Civita tensor $\in_{ijk}$ (Kittel et al., 1962). Putting $\omega_i = \omega_j \delta_{ij}$, where δ_{ij} is the Kronecker symbol, we can modify the angular momentum to read

$$L_i(t) = I_{ij}(t)\,\omega_j(t) + h_i(t) \ , \tag{3.29}$$

where inertial coefficients I_{ij} (time dependent) and functions h_i are defined by

$$I_{ij} = \iiint_V \rho\left(x_k x_k \delta_{ij} - x_i x_j\right)\mathrm{d}V \ ,$$

$$h_i = \iiint_V \rho \in_{ijk} x_j u_k \, \mathrm{d}V \ .$$

The equation of motion (3.25b) can now be modified to read

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\left(I_{ij}\omega_j + h_i\right) + \in_{ijk} \omega_j \left(I_{kl}\omega_l + h_k\right) = N_i} \ , \quad (i = 1, 2, 3) \ . \tag{3.30}$$

These are *Liouville's equations*. If I_{ij}, **h** and **N** are known, we can calculate ω. For example, these equations are used to compute the changes of the Earth's rotation which are caused by the motions of atmospheric or oceanic masses.

As a special case of Liouville's equations we can derive the general equations for a rigid body. Putting $h_i = 0$ and $I_{ij} = \text{const.}$, we arrive at

$$I_{ij}\frac{\mathrm{d}\omega_j}{\mathrm{d}t} + \in_{ijk} \omega_j I_{kl}\omega_l = N_i \ . \tag{3.31}$$

These equations can be simplified further if the coordinate axes coincide with the principal axes of inertia. The tensor of inertia then takes the diagonal form

$$I = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} \ , \tag{3.32}$$

and equations (3.31) reduce to Euler's dynamic equations:

$$A\frac{\mathrm{d}\omega_1}{\mathrm{d}t} - (B - C)\omega_2\omega_3 = N_1 \ ,$$

$$B\frac{\mathrm{d}\omega_2}{\mathrm{d}t}-(C-A)\omega_3\omega_1 = N_2 \ , \tag{3.33}$$

$$C\frac{\mathrm{d}\omega_3}{\mathrm{d}t}-(A-B)\omega_1\omega_2 = N_3 \ .$$

By replacing subscripts 1, 2, 3 with x, y, z, respectively, we arrive at Euler's equations in the form of (3.28).

Chapter 4

Motions of the Earth - Part II

4.1 Time Variations of the Vector of the Angular Velocity of the Earth's Rotation

Astronomical observations carried out from the surface of our planet are affected by the Earth's rotation. At first sight, the rotation seems to be a simple phenomenon, but the increasing accuracy of astronomical observations has revealed the complexity of this motion (Burša and Pec, 1993; Vondrák,1983).

The secular changes of the direction of the Earth's rotation axis in space, although relatively slow, are of a considerable magnitude. As early as in the second century B.C., the Greek astronomer *Hipparchos* discovered *precession*, although this phenomenon was correctly interpreted by Kopernik much later (16^{th} century). Periodic changes in the position of the rotation axis in space, called *nutation*, were observed by *Bradley* in the middle of the 18^{th} century. Theoretically, precession was explained by *Newton* as an effect of the attractive forces of the Sun and Moon on a spherically non-symmetric Earth. D'Alembert then published the first analytical theory, which also explained nutation, in 1749.

That the rotation axis could change its position in the Earth's body (the Earth's poles can move), was first predicted theoretically by *Euler*. From the first half of the 19^{th} century, attempts began to determine such motions from astronomical observations. In 1895 the International Latitude Service (ILS) was founded, and regular observations of latitude variations at several stations of approximately the same latitude, 39°N, began in 1899. In 1962, when the number of these stations had increased, it was replaced by the International Polar Motion Service (IPMS), whose headquarters were in Japan.

In 1911 the International Time Bureau (Bureau International de l'Heure, BIH) was founded in Paris. Its original task was to maintain the international time scale, based on the Earth's rotation. However, soon afterwards it was found hat the rotation speed (absolute value of ω) could not be considered a constant either. This followed from the discrepancies between the Earth's rotation and ephemeris time, which is defined by the motions of the bodies of the Solar System. The situation changed considerably in the 1950's when the first atomic clocks were constructed. The international uniform time scale is now defined by a group of atomic clocks, and astronomical observations serve to determine the irregularities in the rotation of the Earth.

To unite these measurements, the International Earth Rotation Service (IERS) was founded in 1988. The main measuring techniques of this service are as follows:

- VLBI- very long baseline interferometry;
- LLR - lunar laser ranging;

- SLR - satellite laser ranging.

The uniform time is taken from atomic clocks.

The time variations of the vector of the angular velocity of rotation, ω, can be divided as follows:

- variations in direction:
  - – precession and nutation;
  - – the Earth's wobble (also called the motion of the poles, polar wobble, Chandler wobble, or free nutation);
- variations in magnitude (changes of $|\omega|$, changes of the length of day):
  - – deceleration due to tidal friction;
  - – fluctuations;
  - – seasonal variations.

4.2 Precession and Nutation

The Earth's axis of rotation is inclined to the axis of the ecliptic at angle $\in = 23°26'21''$, approximately $23\frac{1}{2}°$. The rotation axis is not fixed in space, but its direction slowly changes. It describes a conic surface about the axis of the ecliptic, with constant angle $\in$ between these axes. This motion of the rotation axis is called *precession* (Fig. 4.1).

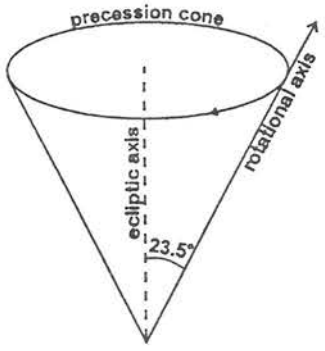

Figure 4.1: Precession of the Earth's rotation axis.

As mentioned above, this motion had already been discovered in the 2nd century B.C. It causes the *motion* of the *celestial pole* and the motion of the *vernal point*. The observed angular velocity of precession is 50.291″ per year. The period of precession is 25 770 years (Platonian year). If seen from the "northern" ecliptic pole, the vector of the angular velocity of the Earth's rotation moves in the clockwise direction (Fig. 4.1).

Precession is a motion which is similar to the motion of a spinning top under the influence of gravity. The first explanation of the causes of precession was given by Newton. He explained that precession was a consequence of the *gravitational effects* of the *Sun and Moon* (the effect of the planets is smaller) *on the flattened Earth.*

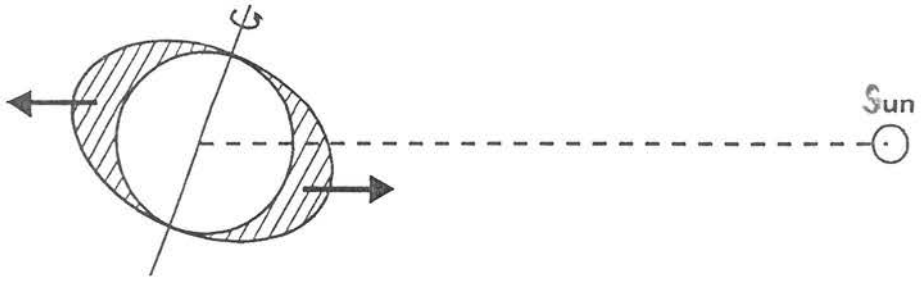

Figure 4.2: Interaction between the Sun and a flattened Earth.

Let us explain the effect of the Sun on the precession of the Earth's rotational axis. In our considerations here (and in all chapters which follow) we may consider the Sun to be a point mass or a spherically symmetric body. Assuming the Earth to be spherically symmetric, the Earth acts on the Sun with a central force, and vice versa. However, the gravitational field of a flattened Earth is non-central, so that the force exerted by the Earth on the Sun is slightly deflected from the Earth-Sun line (Fig. 4.2). In addition to the central component of the force, which is dominant, there is then also a perpendicular component due to the equatorial bulge (tending to pull the Sun closer to the plane of the Earth's equator, i.e. "downward" in Fig. 4.2). The Sun reacts by exerting not only a central force, but also a torque on the flattened Earth (Fig. 4.2). This torque acts about the axis in the equatorial plane which is normal to the Earth-Sun line. It tends to pull the bulge into line with the instantaneous Earth-Sun axis, i.e. to turn the Earth's axis to become perpendicular to the Earth-Sun line. However, this is not generally possible, since the Earth rotates (the Earth's axis is perpendicular to the Earth-Sun line only at the vernal and autumnal equinoxes). Consequently, a precessional motion appears, as with the spinning top.

Actually, there is a small difference between the precession of the Earth and of the top. Namely, the precession of the Earth's axis is not a permanent motion. The precession is maximum at the solstices, when the Sun is at an angular distance of $23\frac{1}{2}^\circ$ from the equatorial plane, and it vanishes at the equinoxes, when the Sun is directly above the equator. But the sense of the precessional motion is the same at both solstices, so that, although it occurs in semi-annual pulses, it has a cumulative effect.

The precession of the Earth's rotational axis, caused by all the bodies of the Solar System, is called *general precession*. It is subdivided into the *luni-solar precession* due to the gravitational effects of the Moon and Sun, which represents the dominant component of the precessional motion, and a much smaller *planetary precession* due to the planets.

Besides the long-period precessional motion, there is a smaller short-period motion of the Earth's axis, called *nutation*. It is caused mainly by the motion of the Moon (as the Moon's orbit does not lie exactly in the plane of the ecliptic). The main nutation term has a period of 18.6 years, its amplitude is $9.206''$. The

superposition of precession and nutation produces an undulatory motion, as if the precessional cone in Fig. 4.1 were corrugated. Actually, there are several nutations, as the orbits are elliptical and inclined.

Theory indicates that the mean angular velocity of precession, caused by the Sun, is

$$\overline{\omega_{p,\,Sun}} = \frac{3}{2}\frac{G}{\omega}\frac{C-A}{C}\frac{M_S}{r_S^3}\cos\epsilon_S \,, \tag{4.1}$$

where G is the gravitational constant, ω is the angular velocity of the Earth's rotation, A is the equatorial and C the polar moment of inertia of the Earth, M_S is the mass of the Sun, r_S is the distance of the Sun from the Earth, ϵ_S is the inclination of the Earth's axis to the ecliptic axis (see Chapter 12). Similarly for the Moon,

$$\overline{\omega_{p,\,Moon}} = \frac{3}{2}\frac{G}{\omega}\frac{C-A}{C}\frac{M_M}{r_M^3}\cos\epsilon_M \,, \tag{4.2}$$

where the subscript M refers to the Moon. As a consequence of term M/r^3, the effect of the Moon is about twice larger than the effect of the Sun; notice the *third power* of the distance in the denominator. The sum of these contributions from the Sun and Moon accounts for the substantial part of the observed precession:

$$\overline{\omega_{p,\,Sun}} + \overline{\omega_{p,\,Moon}} = 50.291'' \text{ per year} \,. \tag{4.3}$$

Since the astronomical parameters in formulae (4.1) and (4.2) are known, we can use formula (4.3) to determine the ratio

$$H = \frac{C-A}{C} \,, \tag{4.4}$$

which is called the *dynamic flattening* of the Earth.

However, formula (4.3) is only approximate, because several factors have not been considered (planetary precession, ellipticities and inclinations of the orbits). A more accurate approach yields

$$H = \left(0.003\,273\,76 \pm 0.000\,000\,02\right) = \frac{1}{305.459 \pm 0.002} \,. \tag{4.5}$$

The geophysical importance of precession is in the following:

1. The measurements of precession make it possible to determine dynamic flattening H. It represents one equation for the determination of the moments of inertia A and C. We shall see that the second equation is

$J_2 = \frac{C-A}{Ma^2} = 1\,082.64 \times 10^{-6}$, as determined from satellite measurements. Consequently, $C = \frac{J_2}{H} Ma^2 = 0.330\,702 Ma^2$. For a homogeneous sphere, $C = 0.4 Ma^2$. Hence, there must be an accumulation of mass towards the Earth's centre. The mass of the Earth and moment of inertia C represent the basic constraints on the density distribution within the Earth.

2. Precession contributes as an energy source to some geophysical processes (geomagnetic dynamo).
3. We have seen that nutation causes variations of the inclination of the Earth's rotational axis, but these variations are relatively small and short-period. However, some components of the planetary precession cause more pronounced variations in the inclination. This is important for the evolution of the climate, e.g., for the occurrence of ice ages. We shall return to this problem below in this chapter.

4.3 Polar Motion

The rotation axis not only changes its direction in space, but its position also changes slightly within the Earth's body. This motion is called *free nutation* (but it is not strictly a nutation), the *wobble* or the *Chandler wobble*. It causes a small wandering of the geographical poles, variations of latitude and variations of the heights of stars. Consequently, the geographical coordinates of points on the Earth's surface vary in time. It must be taken into account in determining the mean geographical coordinates.

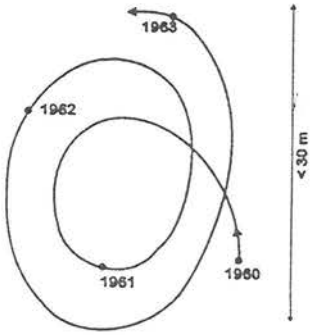

Figure 4.3: Wandering curve of the North Pole. (Simplified from Burša and Pec (1993)).

The pole moves along a complicated curve in the sense of rotation (Fig. 4.3). This curve can be circumscribed by a circle whose radius on the Earth's surface is not larger than about 10 m. In other words, the angle α between the rotation axis and the principal polar axis of inertia is less than $0.3''$.

The motion of the poles is composed of four components:

1. annual component (period of 12 months);
2. *Chandler's component* with a period of about *14 months* (425-440 days, mean value of about 433 days);
3. small semi-annual component;
4. irregular components.

As mentioned above, the theoretical explanation of the polar motion was given by Euler, as the motion of a force-free top. To simplify the corresponding equations, let us assume that:

1. The Earth is a rigid body.
2. The coordinate system, fixed in the Earth, has its origin at the centre of mass and axes x, y, z coincide with the principal axes of inertia, the z-axis being close to the rotation axis.
3. For shorter periods of time (years) we can approximately neglect the external torques caused by the Moon and Sun, i.e. we shall put $N_x = N_y = N_z = 0$. These torques are very small and have some effects only over a longer time interval; see the precession and nutation.
4. The moments of inertia with respect to the principal equatorial axes are the same, $A = B$ (e.g., the Earth is rotationally symmetrical), but the moment of inertia C with respect to the z-axis is different, $C > A$ (as the Earth is flattened at the poles).

Euler's dynamic equations, describing the motion of the rigid body,

$$\begin{aligned} A\dot{\omega}_x - (B-C)\omega_y\omega_z &= N_x \ , \\ B\dot{\omega}_y - (C-A)\omega_z\omega_x &= N_y \ , \\ C\dot{\omega}_z - (A-B)\omega_x\omega_y &= N_z \end{aligned} \qquad (4.6)$$

then simplify to read

$$\begin{aligned} A\dot{\omega}_x - (A-C)\omega_y\omega_z &= 0 \ , \\ A\dot{\omega}_y - (C-A)\omega_z\omega_x &= 0 \ , \\ C\dot{\omega}_z &= 0 \ . \end{aligned} \qquad (4.7)$$

The last equation indicates that $\omega_z = \omega_o = \text{const}$. Differentiating the first equation with respect to time,

$$A\ddot{\omega}_x - (A-C)\dot{\omega}_y\omega_0 = 0$$

and substituting for $\dot{\omega}_y$ from the second equation yields the equation of a harmonic oscillator,

$$\ddot{\omega}_x + \left(\frac{C-A}{A}\right)^2 \omega_0^2 \omega_x = 0 \ . \tag{4.8}$$

Putting

$$\Omega = \frac{C-A}{A}\omega_0 \ , \tag{4.9}$$

the general solution of equation (4.8) can be expressed as

$$\omega_x = ae^{i\Omega t} + be^{-i\Omega t} \ ,$$

or

$$\omega_x = F\cos\left(\Omega t + \gamma\right) \ , \tag{4.10}$$

and similarly

$$\omega_y = F\sin\left(\Omega t + \gamma\right) \ , \tag{4.11}$$

where constants F and γ must be determined from the initial conditions. The period of this motion is

$$T = \frac{2\pi}{\Omega} = \frac{2\pi}{\omega_0}\frac{A}{C-A} \ .$$

However, $\dfrac{2\pi}{\omega_0} = 1$ day, so that

$$\boxed{T_{days} = \frac{A}{C-A} \approx \frac{C}{C-A} = 305 \text{ days}} \ . \tag{4.12}$$

We have just arrived at the following conclusion. If, for some reason, the rotation axis does not coincide with a principal axis of inertia, it must rotate about the principal axis (Fig. 4.4).

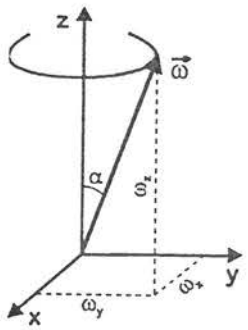

Figure 4.4: Free nutation of a rigid Earth.

We have assumed that the Earth was rigid. However, the observed period is longer, equal to *Chandler's period* (433 days). This discrepancy was explained by Newcomb (1891) mainly as a consequence of the elasticity of the Earth. Thus, Chandler's period can be used to obtain some information on the elasticity of the Earth. (More specifically, on Love number k, see the next chapter.) Theoretical periods calculated for several simplified models are given in Tab. 4.1.

Table 4.1. Theoretical periods of the polar motion.

Model	Period (sidereal days)
Rigid Earth	305
Rigid mantle, liquid core	272 (Kelvin, 1876)
Elastic solid Earth	319
Elastic mantle, liquid core	401

The difference of the last value in Tab. 4.1 from Chandler's period (about 30 days) seems to be caused by the effect of the ocean, which was not considered in the computations. The table also indicates that the existence of a liquid core could have theoretically been predicted at the end of the 19th century, before the seismological evidence appeared.

Finally, two remarks should be added:

1. The Chandler wobble is *damped*, with a damping time of around 30 years (10-70 years). Consequently, after some time, angle α would approach to zero. Therefore, there must be processes that excite the Chandler wobble (earthquakes, meteorological or hydrological processes). The problem of the *excitation of the Chandler wobble* still remains open.
2. *Euler's equations* represent a system of *non-linear* differential equations. Assuming $\mathbf{N} = 0$ and $A = B$, the system reduced to two linear equations. For $\mathbf{N} = 0$ and $A \neq B$, the system remains non-linear, but the solution can still be expressed in analytical form in terms of elliptical functions.

4.4 Other Long-Period Motions

There are also small variations in the orbital parameters of the Earth and in the inclination of its axis (Koudelková, 1993):

- eccentricity e varies from 0 up to 6 %; period ~ 100 000 years; (now $e = 1.7$ %);
- inclination $\in = 23°26' \pm 1°30'$; period ~ 41 000 years; (now $\in = 23°26'$).

The time variation of the inclination $\in$ is caused by the inclination of the orbital planes of the other planets to the orbital plane of the Earth. (As the inclination of the Moon's orbit causes nutation, these inclinations produce additional contributions.)

These two motions play an important role in the *changes of the climate*, in the occurrence of *ice ages* (Milankovich's hypothesis). Meteorological

modellings show that the development of the climate is controlled more by the meteorological situation in summer than in winter. Consequently:

large $\in$ $\Rightarrow$ hot summer (and cold winters) $\Rightarrow$ warmer climate;

small $\in$ $\Rightarrow$ colder summers $\Rightarrow$ *ice ages.*

It seems that this simplified model can explain the main features of the last Ice Age (Tab. 4.2):

Table 4.2. Correlation between the changes of the inclination of the Earth's axis and climatic changes.

Time (years)	Inclination	Insolation (compared with the present)	Beginning and duration of climatic epochs
127 000 B.P.			interglacial
125 000 B.P.	23°48′	+13 %	(12 000 years)
115 000 B.P.	22°24′	–9 %	Ice Age
18 000 B.P.	culmination of the Ice Age		(~100 000 years)
11 000 B.P.	24°12′	+10 %	present
0 (now)	23°26′	0 %	interglacial
5 000 A.P.			next Ice Age
70 000 A.P.			next interglacial

4.5 Changes of the Length of Day

Several components can be distinguished in the changes of the magnitude of the angular velocity of rotation, i.e. of $|\omega|$:

1. *Secular deceleration* of the rotation, caused by *tidal friction.* This deceleration causes a prolongation of the day by 0.001 7 seconds per 100 years, i.e. the prolongation of the day by *1 sec per 60 000 years.* It was determined from astronomical data (eclipses) and from the growth of corals. For example, in the Devonian period (~ 380 million years ago) a year had about 400 days.
2. *Fluctuations*, which represent irregular variations of the Earth' rotation with periods of about 10-100 years. The fluctuations are probably caused by geophysical phenomena (by motions within the Earth, which lead to changes of the moments of inertia and, consequently, to changes of ω). For example, it is known that the Earth's mantle and core rotate at different velocities, but certain mechanical and electromagnetic couplings exist between them. The changes in these couplings must then lead to changes of the length of day.

3. *Seasonal variations*, which are caused by meteorological factors (Figs. 4.5 and 4.6). In March, the angular velocity of rotation ω has a minimum (the slowest rotation), and in July - August it has a maximum (the fastest rotation). Moreover, a secondary minimum of ω occurs in November, and a secondary maximum in January.

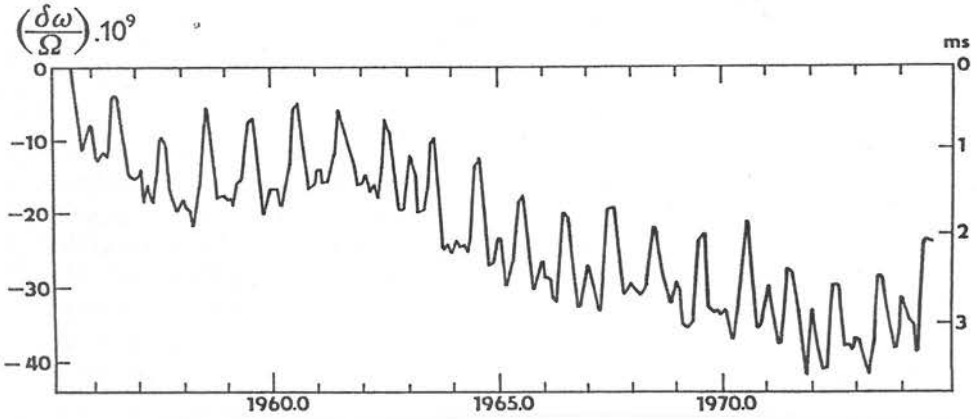

Figure 4.5: Variations in the Earth's angular velocity of rotation ω (left-hand scale) and in the length of day T (right-hand scale); $\Omega = 2\pi/86\,400$ s. (After Burša and Pec (1993)).

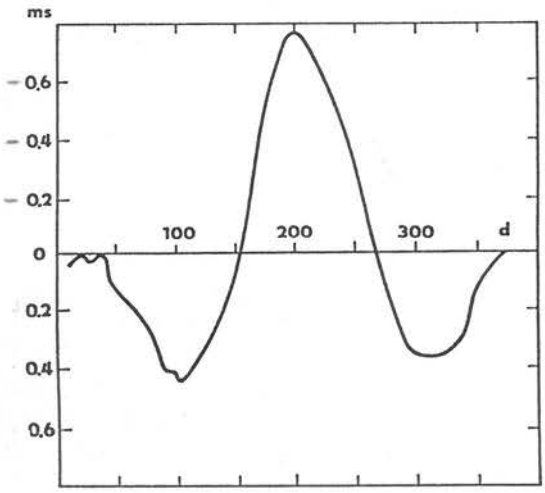

Figure 4.6: Mean annual curve of seasonal diurnal variations in the length of day calculated for the period 1962–1978. (Modified from Burša and Pec (1993)).

4.6 Dynamics of the Earth-Moon System

A more detailed theory should consider the Earth-Moon-Sun system, but the effect of the Sun is smaller and, therefore, it will be neglected here.

Let us look at the Earth-Moon system from the northern side of the Moon's orbital plane (Fig. 4.7). Let us neglect the inclination of the Earth's rotational

axis, and assume simply that it is perpendicular to the orbital plane of the Moon. Let us introduce the following notations: T is the centre of mass of the system, ω the angular velocity of the Earth's rotation, ω_L the angular velocity of the orbital motion of the Earth and Moon about the common centre of mass T, R the distance between the centres of the bodies, δ the lag of the tidal bulge of the Earth relative to the Earth-Moon axis, M the mass of the Earth, m the mass of the Moon, and C the moment of inertia of the Earth.

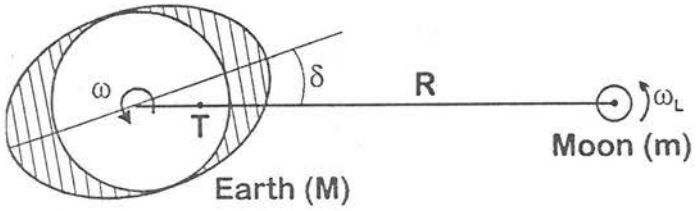

Figure 4.7: Situation characterising the dynamics of the Earth-Moon system.

In an inertial system, the Earth rotates from west to east, and the Moon moves in the same direction. However, as $\omega > \omega_L$, it seems that the Moon revolves round the Earth from east to west. As a result of energy dissipation, the tidal deformation is not in phase with the position of the Moon (see lag angle δ). The Moon exerts a larger force on the nearer bulge than on the more distant one. It produces a torque which tries to slow the Earth's rotation. As a reaction, a force is generated which pulls the Moon in the direction of its motion. Thus, forces are generated which try to decelerate the Earth's rotation and to accelerate the Moon's orbital motion. But any acceleration of a satellite in the direction of its motion moves the satellite into a higher orbit where the orbital angular velocity is smaller. Consequently, the lag of the tidal bulge causes a decrease of ω, increase of R, and (surprisingly) also a decrease of ω_L.

4.6.1 Basic numerical values

The following astronomical parameters, characterising the Earth, Moon and their motions, will be used below (Burša & Pec, 1993; Burša et al., 1995):

$$GM = 398\,600.44 \times 10^9 \text{ m}^3\text{s}^{-2},$$
$$Gm = 4\,902.80 \times 10^9 \text{ m}^3\text{s}^{-2},$$
$$\omega = 7.292\,115 \times 10^{-5} \text{ rad s}^{-1},$$
$$\omega_L = 2.661\,9 \times 10^{-6} \text{ rad s}^{-1},$$
$$R = 384\,400 \times 10^3 \text{ m},$$
$$C = 8.036\,5 \times 10^{37} \text{ kg m}^2.$$

The observed data, concerning the dynamics of the system, are as follows:
a) From the growth of corals,

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -(5.4 \pm 0.5) \times 10^{-22} \text{ rad s}^{-2} \ . \tag{4.13}$$

b) From astronomical data (last 2 700 years),

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -(4.5 \pm 0.1) \times 10^{-22} \text{ rad s}^{-2} \ . \tag{4.14}$$

c) From Lunar Laser Ranging (LLR, direct measurements),

$$\frac{\mathrm{d}\omega_L}{\mathrm{d}t} = -(25.88 \pm 0.5)'' \text{ century}^{-2} = -1.260 \times 10^{-23} \text{ rad s}^{-2} \ , \tag{4.15}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = (3.77 \pm 0.15) \text{ cm/year} = 1.19 \times 10^{-9} \text{ m s}^{-1} \ . \tag{4.16}$$

d) From perturbations of satellite orbits,

$$\delta = 2.9° \ . \tag{4.17}$$

4.6.2 Basic theory and equations

The basic theory of the Earth-Moon dynamics is based on the following principles and assumptions:

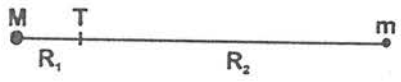

Figure 4.8: Position of the centre of mass of a two-body system.

1. The *gravitational and centrifugal* forces acting on the Moon have the *same magnitude*.

   Denote R_1 and R_2 the distances of the centres of the Earth and Moon, respectively, from the common centre of mass, T (Fig. 4.8). It is well known that

$$R_1 = \frac{m}{M+m} R \ , \quad R_2 = \frac{M}{M+m} R \ , \tag{4.18}$$

   where $R = R_1 + R_2$. The balance between the centrifugal and gravitational forces can then be expressed as

$$m\omega_L^2 R_2 = \frac{GMm}{R^2} \ . \tag{4.19}$$

Using (4.18) yields Kepler's third law (for a circular orbit),

$$\omega_L^2 R^3 = G(M+m) \ . \tag{4.20}$$

By differentiating this equation with respect to time we obtain

$$2\omega_L \frac{\mathrm{d}\,\omega_L}{\mathrm{d}\,t} R^3 + 3\omega_L^2 R^2 \frac{\mathrm{d}\,R}{\mathrm{d}\,t} = 0 \ ,$$

which yields

$$\frac{\mathrm{d}\,R}{\mathrm{d}\,t} = -\frac{2}{3}\frac{R}{\omega_L}\frac{\mathrm{d}\,\omega_L}{\mathrm{d}\,t} = 3.83 \text{ cm/year} \ . \tag{4.21}$$

This numerical value is in good agreement with value (4.16), which was determined independently. Extrapolating this value back in time to the origin of the Earth (4.5 billion years ago), we obtain a total change of 170 000 km, i.e. half the present distance between the bodies. Although this extrapolation is not fully justified, it indicates that tides and tidal friction have played an important role in the evolution of the Earth and Moon.

2. *Conservation of the total angular momentum.*
As we neglect the external torques (especially that exerted by the Sun), the total angular momentum of the system remains constant. Neglecting also the small angular momentum of the Moon's rotation $\left(= C_L\omega_L\right)$, we have

$$\underbrace{C\omega}_{\substack{\text{rotational}\\ \text{of the Earth}}} + \underbrace{R_1 M v_1}_{\substack{\text{orbital}\\ \text{of the Earth}}} + \underbrace{R_2 m v_2}_{\substack{\text{orbital}\\ \text{of the Moon}}} = \text{const} \ , \tag{4.22}$$

where v_1 and v_2 are the orbital velocities of the Earth and Moon, respectively. As $v_1 = R_1\omega_L$ and $v_2 = R_2\omega_L$, it follows from (4.18) and (4.22) that

$$C\omega + \omega_L M\left(\frac{m}{M+m}R\right)^2 + \omega_L m\left(\frac{M}{M+m}R\right)^2 = \text{const} \ ,$$

i.e.

$$C\omega + \omega_L R^2 \frac{Mm}{M+m} = \text{const} \ . \tag{4.23}$$

By differentiating this equation with respect to time and using (4.21) we obtain

$$\frac{\mathrm{d}(C\omega)}{\mathrm{d}t} = \frac{1}{3}R^2 \frac{Mm}{M+m}\frac{\mathrm{d}\omega_L}{\mathrm{d}t} = -4.5\times10^{16}\,\mathrm{kg\,m^2\,s^{-2}} \ . \qquad (4.24)$$

Assuming C to be constant yields

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -5.6\times10^{-22}\,\mathrm{rad\,s^{-2}} \ . \qquad (4.25)$$

This value follows from the tidal dynamics of the Earth-Moon system, assuming C = const. It is close to (4.13) but differs from (4.14).

If a more accurate theory of the Earth-Moon-Sun system is used, instead of (4.25) we arrive at a higher value:

$$\left(\frac{\mathrm{d}\omega}{\mathrm{d}t}\right)_{tidal} = (-6.1\pm0.4)\times10^{-22}\,\mathrm{rad\,s^{-2}} \ . \qquad (4.26)$$

Considering (4.14) as the most probable value of the Earth's rotation deceleration, one can see that there is an accelerating process which partly (to 1/4) compensates tidal deceleration (4.26):

$$\left(\frac{\mathrm{d}\omega}{\mathrm{d}t}\right)_{non-tidal} = (+1.6\pm0.4)\times10^{-22}\,\mathrm{rad\,s^{-2}} \ . \qquad (4.27)$$

The corresponding change of the moment of inertia satisfies the equation

$$C\left(\frac{\mathrm{d}\omega}{\mathrm{d}t}\right)_{non-tidal} + \frac{\mathrm{d}C}{\mathrm{d}t}\omega = 0 \ ,$$

which yields

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -1.8\times10^{20}\,\mathrm{kg\,m^2\,s^{-1}} \ . \qquad (4.28)$$

It indicates that a transport of mass toward the Earth's centre must exist. Various processes have been considered, such as melting of glaciers or internal processes. Nevertheless, the problem remains open; the magnitude of $\mathrm{d}C/\mathrm{d}t$ is too large to be explained by a simple process. The decrease of C with time is also in sharp *contradiction* with hypotheses on the *global expansion* of the Earth; in an expanding Earth the moment of inertia should increase, not decrease.

The decrease of C with time also seems to be supported by the measurements of the change of the second-degree zonal geopotential (Stokes) parameter

$$J_2 = \frac{C - \dfrac{A+B}{2}}{Ma^2}\,, \tag{4.29}$$

where a is the equatorial radius of the Earth. The long-term variation in J_2, determined from satellite motions (LAGEOS and others), is

$$\frac{\mathrm{d}J_2}{\mathrm{d}t} = -(2.6 \pm 0.3) \times 10^{-9}\,\text{century}^{-1}\,. \tag{4.30}$$

3. The *mechanical energy* of the system is *not conserved.*

Dissipation processes transform a part of the mechanical energy into heat. Neglecting the kinetic energy of the Moon's rotation about its axis, the mechanical energy of the Earth-Moon system is

$$E = \frac{1}{2}C\omega^2 + \frac{1}{2}Mv_1^2 + \frac{1}{2}mv_2^2 - \frac{GMm}{R}\,, \tag{4.31}$$

where the individual terms on the right-hand side represent the rotational kinetic energy of the Earth, orbital kinetic energy of the Earth, orbital kinetic energy of the Moon, and the potential energy of the system, respectively. From analogies with (4.22) and (4.23) we arrive at

$$E = \frac{1}{2}C\omega^2 + \frac{1}{2}\omega_L^2 R^2 \frac{Mm}{M+m} - \frac{GMm}{R}\,. \tag{4.32}$$

Substituting Kepler's third law (4.20) into the second term on the right-hand side yields

$$E = \frac{1}{2}C\omega^2 - \frac{1}{2}\frac{GMm}{R}\,. \tag{4.33}$$

The tidal dissipation of the mechanical energy is given by the derivative

$$\frac{\mathrm{d}E}{\mathrm{d}t} = C\omega\frac{\mathrm{d}\omega}{\mathrm{d}t} + \frac{1}{2}\frac{GMm}{R^2}\frac{\mathrm{d}R}{\mathrm{d}t}\,. \tag{4.34}$$

However, using (4.21), (4.24) and (4.20) produces

$$\frac{\mathrm{d}R}{\mathrm{d}t} = -\frac{2}{3}\frac{R}{\omega_L}\frac{\mathrm{d}\omega_L}{\mathrm{d}t} = -\frac{2}{3}\frac{R}{\omega_L}C\frac{\mathrm{d}\omega}{\mathrm{d}t}\frac{3(M+m)}{R^2 Mm} = -2\omega_L R^2 C\frac{\mathrm{d}\omega}{\mathrm{d}t}\frac{1}{GMm}\,.$$

Consequently, formula (4.34) takes a very simple form:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = C\left(\omega - \omega_L\right)\frac{\mathrm{d}\omega}{\mathrm{d}t} \ . \tag{4.35}$$

Substituting (4.25) for the tidal deceleration in the Earth-Moon system, we arrive at

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -3.2 \times 10^{12}\ \mathrm{W} \ . \tag{4.36}$$

The higher deceleration given by (4.26) yields

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -3.4 \times 10^{12}\ \mathrm{W} \ . \tag{4.37}$$

This power may be important in generating the *geomagnetic field* and *heat flow* from the Earth's interior. It contributes to these phenomena.

In conclusion it should be re-iterated that we have only just started with accurate measurements of long-term changes of the Earth's rotation. Nevertheless, it is becoming evident that these investigations may throw a new light on many fundamental problems of the evolution and dynamics of the Earth.

Chapter 5

Earth Tides

Marine tides were already known in antiquity, but their scientific explanation was only given by *Newton*. He explained them as the effect of the gravitational attraction of the Moon and Sun on the oceans. The largest difference between the height of the high tide and low tide occurs in the Bay of Fundy, on the eastern coast of Canada. The difference there is as much as 50 feet, but this is the result of a resonance amplification in the coastal basin. In the open oceans the difference is only a few feet.

However, there are also the tides of the solid parts of the Earth, called *earth tides*, which represent the deformations of the solid Earth caused by the gravitational fields of the Moon and Sun. The effect of the other celestial bodies can be neglected. The tides cause specific motions of the Earth which were not considered in the previous chapters.

There are several reasons why tides are studied in geophysics:

1. Tides cause periodic variations of the gravity field. Consequently, these variations must be removed from accurate gravimetric measurements, i.e. *tidal corrections* must be introduced.
2. Tides can be used to study some *physical parameters* of the Earth as a whole, especially certain elastic parameters of the Earth. If the Earth were rigid, the gravity variations could be calculated from the known positions of the Moon and Sun. However, the Earth deforms under the influence of tidal forces, so that the measured gravity variation slightly exceeds the theoretical values calculated for a rigid Earth. This degree of amplification characterises the elasticity of the Earth as a whole.
3. Tides contribute to some geophysical processes as *sources of energy*. We have already mentioned the tidal power of about 3×10^{12} W dissipated in the Earth. Tides play a certain role in tectonic processes and in generating the geomagnetic field.

In this chapter we shall present only the basic theory of the tidal phenomena. More comprehensive treatments can be found in Melchior (1966, 1983), and Pick et al. (1973).

5.1 Tidal Effects on a Rigid Earth

5.1.1 Origin of tidal forces

In this section we shall not consider tidal deformations of the Earth's body yet, but only deformations of equipotential surfaces. Let us consider the tides due to the Sun; the Moon will be taken into account later. Thus, in order to simplify the problem, let us adopt the following assumptions:

1. The Earth is *rigid*.

2. The Earth revolves round the Sun along a *circular orbit*.
3. The Earth *does not rotate* about its axis. This means, for example, that any bar fixed to the Earth preserves its direction with respect to the inertial reference frame (see the motion of point A with respect to the Earth's centre in Fig. 5.1).

Assumptions 1 and 2 are relatively reasonable, but the third assumption contradicts reality. We shall abandon this assumption later.

As a consequence of assumptions 2 and 3, all points of the Earth move along circular orbits of the same radius, so that the centrifugal acceleration has the same magnitude at all points (Fig. 5.1). Moreover, at a given time, centrifugal accelerations are parallel at all points of the Earth. It means that at a given time the *centrifugal acceleration* forms a *homogeneous field.*

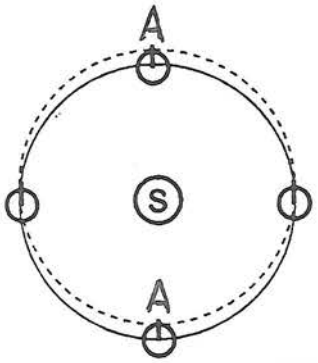

Figure 5.1: Revolution of a non-rotating Earth around the Sun. The solid line shows the orbit of the Earth's centre, the dashed line is the orbit of another point A.

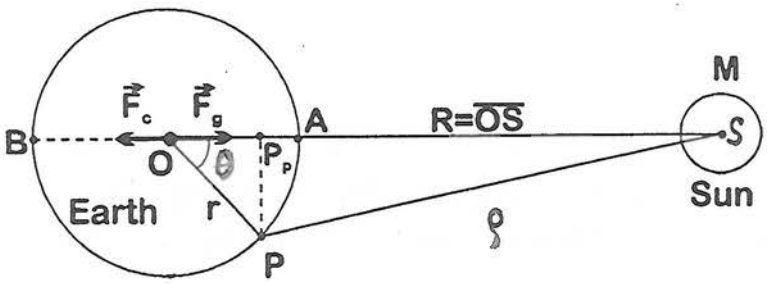

Figure 5.2: Geometry for calculating the tidal potential of the Sun.

The situation shown in Fig. 5.1 is enlarged in Fig. 5.2 (the proportions again are not preserved). Compare the forces acting at the following points of the Earth: at the nearest point to the Sun, A, at the Earth's centre O, and at the most distant point B. (Actually we shall not compare forces, but intensities, and they are identical with accelerations in these parts of mechanics). As we have shown, the centrifugal acceleration, F_c, is the same at all points of the Earth, which means that $F_c = F_c(A) = F_c(O) = F_c(B)$. However, the gravitational

acceleration, F_g, decreases from A to B. Both accelerations are balanced at the Earth's centre O, since we have assumed a circular orbit of the Earth about the Sun. Consequently,

at O: $F_g = F_c$;

at A: $F_g(A) > F_g(O) = F_c(O) = F_c(A)$, i.e. $F_g > F_c$;

at B: $F_g < F_c$.

This causes the deformation of the equipotential surface as shown in Fig. 5.3. Tidal bulges appear on both sides of the Earth (on the near side and on the far side). Consequently, on the rotating Earth, the high tide and the low tide occur twice a day (two ebbs and two flows).

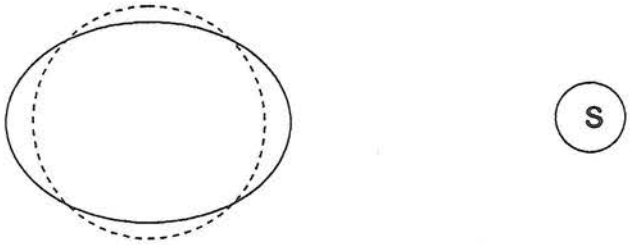

Figure 5.3: Tidal deformation of the equipotential surface (solid line).

As the gravitational attraction of the Sun is a *central* force and the centrifugal force forms a *homogeneous* force field, they cannot compensate each other completely. The resultant force, the so-called *tidal force*, or tide-generating force, is generally non-zero. The intensity of the tidal force (tidal acceleration) is

$$\mathbf{F}_t = \mathbf{F}_g + \mathbf{F}_c \ . \tag{5.1}$$

We shall first derive the expressions for the corresponding potentials.

5.1.2 Tidal potential

In physics it is usual to define potential V as a scalar function which is related to intensity $\mathbf{F}$ as

$$\mathbf{F} = -\mathrm{grad}\, V \ . \tag{5.2}$$

The potential V_g of the gravitational field of the Sun at point P is then

$$V_g(P) = -\frac{GM}{\rho} \ , \tag{5.3}$$

where G is the gravitational constant, M the mass of the Sun, and ρ the distance of point P from the centre of the Sun, S (Fig. 5.2).

The centrifugal intensity relates to the gravitational intensity as

$$F_c = F_g(O) = \frac{GM}{R^2}\,, \tag{5.4}$$

where $R = \overline{OS}$ is the distance between the centres of the Earth and Sun. Intensity $\mathbf{F}_c$ at all points of the Earth is parallel to the instantaneous Earth-Sun axis OS. Thus, the equipotential surfaces of the centrifugal potential (at a given moment) are the planes which are perpendicular to the OS axis. Let us denote this axis as the z-axis, put its origin at the Earth's centre O, and orient it positively toward the Sun. Let us select the equipotential plane passing through the Earth's centre O, i.e. plane $z = 0$, as the surface of the zero centrifugal potential. Centrifugal potential V_c at point P will then be equal, since the field is homogeneous, to product $\pm F_c \cdot z$; here F_c is the magnitude of the centrifugal intensity, given by (5.4), and z is the coordinate of the point just introduced, which characterises the distance of the point from the plane of the zero potential. It remains to select the correct sign of the product. According to (5.2), the intensity is oriented in the direction of the steepest descent of the potential. Therefore, since intensity $\mathbf{F}_c$ is directed in the negative sense of the z-axis, potential V_c must increase in the opposite direction, i.e. the positive sign is appropriate. Moreover, let us express Cartesian coordinate z in terms of spherical coordinates r and θ, where r is the radial distance from the Earth's centre, and θ is the angular distance from the z-axis. Finally,

$$V_c(P) = F_c \cdot z = \frac{GM}{R^2} r\cos\theta\,. \tag{5.5}$$

The tidal potential is then

$$V_t = V_g + V_c + const\,, \tag{5.6}$$

i.e.

$$V_t(P) = -\frac{GM}{\rho} + \frac{GM}{R^2} r\cos\theta + const\,. \tag{5.7}$$

Select the arbitrary constant that the tidal potential at the Earth's centre is zero, $V_t(O) = 0$:

$$0 = V_t(O) = -\frac{GM}{R} + 0 + const \quad \Rightarrow \quad const = \frac{GM}{R}\,.$$

Consequently,

$$V_t(P) = -\frac{GM}{\rho} + \frac{GM}{R^2} r\cos\theta + \frac{GM}{R} = -GM\left(\frac{1}{\rho} - \frac{1}{R} - \frac{r\cos\theta}{R^2}\right)\,. \tag{5.8}$$

We would like to express distance ρ in terms of the parameters which describe the position of the perturbing body and the point of observation. Applying the cosine theorem to triangle *OPS*, we arrive at

$$\rho^2 = R^2 - 2rR\cos\theta + r^2 ,$$

$$\frac{1}{\rho} = \frac{1}{R}\frac{1}{\sqrt{1-2\frac{r}{R}\cos\theta+\left(\frac{r}{R}\right)^2}} . \tag{5.9}$$

Put $u = r/R$ and

$$f\left(u,\cos\theta\right) = \left(1-2u\cos\theta+u^2\right)^{-\frac{1}{2}} . \tag{5.10}$$

Function f is a function of two variables, u and θ (or $\cos\theta$). Since u is a small parameter, it will be sufficient to approximate function f by a few terms of its Taylor series in u in the vicinity of $u = 0$. Keeping parameter θ fixed for this moment, we can put

$$f(u) = f(0) + \frac{f'(0)}{1!}u + \frac{f''(0)}{2!}u^2 + \cdots , \tag{5.11}$$

where f', f'', etc., represent the partial derivatives of f with respect to u. It follows from (5.10) that

$$\begin{aligned}
&f(0) = 1 ,\\
&f'(u) = -\frac{1}{2}\left(1-2u\cos\theta+u^2\right)^{-\frac{3}{2}}(-2\cos\theta+2u), \quad f'(0) = \cos\theta ,\\
&f''(u) = -\frac{1}{2}\left(-\frac{3}{2}\right)\left(1-2u\cos\theta+u^2\right)^{-\frac{5}{2}}(-2\cos\theta+2u)^2 -\\
&\qquad -\frac{1}{2}\left(1-2u\cos\theta+u^2\right)^{-\frac{3}{2}}\cdot 2 , \quad f''(0) = 3\cos^2\theta - 1 .
\end{aligned} \tag{5.12}$$

Consequently, the reciprocal distance $1/\rho$, see (5.9), can be expressed as

$$\frac{1}{\rho} = \frac{1}{R}\left[1+\frac{r}{R}\cos\theta+\frac{1}{2}\frac{r^2}{R^2}\left(3\cos^2\theta-1\right)+\cdots\right] . \tag{5.13}$$

By substituting this series into tidal potential (5.8), the first two terms of the series vanish. Retaining only the first non-zero term yields

$$V_t(P) = -\frac{GMr^2}{2R^3}\left(3\cos^2\theta - 1\right) . \tag{5.14}$$

This formula is usually modified by introducing the double angle 2θ. Using

$$\cos^2\theta = \frac{1+\cos 2\theta}{2} ,$$

we finally arrive at

$$\boxed{V_t(P) = -\frac{GMr^2}{4R^3}(3\cos 2\theta + 1)} . \tag{5.15}$$

Before analysing the properties of the tidal potential, note that in the next chapter we shall express series (5.11) in the form

$$f(u) = \sum_{n=0}^{\infty} u^n P_n(\cos\theta) , \tag{5.16}$$

where $P_n(\cos\theta)$ are Legendre polynomials (of variable $\cos\theta$). Equations (5.11) and (5.12) indicate that

$$P_0(\cos\theta) = 1, \quad P_1(\cos\theta) = \cos\theta, \quad P_2(\cos\theta) = \frac{3}{2}\cos^2\theta - \frac{1}{2} . \tag{5.17}$$

In mathematics, astronomy, geodesy and geophysics we often introduce potentials with opposite signs than in physics. We shall also do this here. Changing the sign in (5.15) produces

$$V_t(P) = +\frac{GMr^2}{4R^3}(3\cos 2\theta + 1) , \tag{5.18}$$

but to preserve the correct direction of intensity we must express the tidal intensity as

$$\mathbf{F}_t = +\mathrm{grad}\, V_t . \tag{5.19}$$

The first term (5.18) of the tidal series yields the tides which are symmetrical on both sides of the Earth. The next term, which is usually omitted, would provide a small asymmetric contribution.

Using subscript S for the parameters of the Sun, we shall express the solar tidal potential (5.18) as follows:

$$\boxed{V_{t,S} = \frac{GM_S r^2}{4R_S^3}\left(3\cos 2\theta_S + 1\right)} . \tag{5.20}$$

Analogously, for the tidal potential due to the Moon,

$$\boxed{V_{t,M} = \frac{GM_M r^2}{4R_M^3}\left(3\cos 2\theta_M + 1\right)}, \tag{5.21}$$

where subscript M denotes the Moon.

Let us add an explanation concerning the latter formula. We should expect the formula for the lunar tides to be similar to that for the solar tides. On the other hand, the physical situation seems to be a little different, because we have assumed that the Earth revolves round the Sun, but we cannot adopt the assumption that the Earth revolves round the Moon. However, to be stricter, we should say that the Earth does not revolve round the centre of the Sun, as shown in Fig. 5.1, but that both the bodies revolve round a common centre of mass. Formula (5.4) for the centrifugal acceleration need not be changed; it is correct, it describes this very situation. (In other words, although Fig. 5.1 and our descriptions were not accurate, the basic formula (5.4) was correct). Analogously, the Earth and Moon revolve round a common centre of mass, which substantiates the validity of formula (5.18) also for the lunar tides.

5.1.3 Properties of the tidal potential

Let us discuss some properties of the tidal potential (5.18). Firstly, we should mention the dependence on the double angle, 2θ. This indicates the symmetry of the tides, given by the leading term (5.18), with respect to plane $\theta = 90°$. Consequently, the tides are the same on the near side and the far side with respect to the perturbing body (the Sun or Moon). On the rotating Earth it produces two high tides and two low tides a day.

Secondly, tidal potential (5.18) changes its sign for $3\cos 2\theta + 1 = 0$, i.e. for θ of about 55° (a more accurate value is $\theta = 54.7°$). For $\theta < 55°$ the potential is positive, which yields the high tide (on a spherical cap with an apex angle of 109°). In the belt $55° < \theta < 125°$ the tide is low, and for $125° < \theta \leq 180°$ the tide is high again (Fig 5.4).

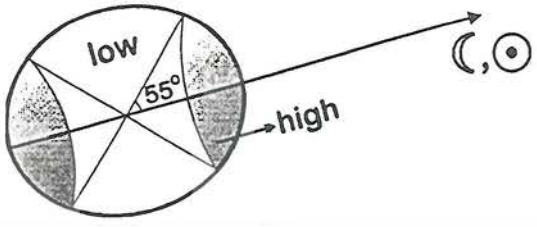

Figure 5.4: Regions of high tide (hatched caps) and of low tide (white belt).

Thirdly, the parentheses in (5.18) indicate that tides are not symmetrical with respect to their maximum and minimum values, because the values of $3\cos 2\theta + 1$ vary between +4 and -2.

Fourthly, the tidal potential decreases with the third power of distance of the perturbing body, i.e. as $1/R^3$. Realising that the potential of a point source decreases as $1/R$, and the potential of a dipole as $1/R^2$, we see that the tidal potential has the character of the potential of a quadrupole. Despite the large mass of the Sun, it is responsible for the solar tides being about twice smaller than the lunar tides. Considering the following values (Burša et al., 1995),

$$\begin{aligned} GM_S &= 13\,271\,244.0 \times 10^{13}\,\mathrm{m^3\ s^{-2}}\ , \\ R_S &= 1\,\mathrm{AU} = 1.495\,978\,7 \times 10^{11}\,\mathrm{m}\ , \\ GM_M &= 4\,902.799 \times 10^{9}\,\mathrm{m^3\ s^{-2}}\ , \\ R_M &= 384\,400\,\mathrm{km}\ , \end{aligned} \tag{5.22}$$

we arrive at the ratio

$$\frac{M_S}{R_S^3} = 0.46\frac{M_M}{R_M^3}\ . \tag{5.23}$$

Consequently, the lunar tides are slightly more than twice higher than the solar tides. Note that the same factor, M/R^3, also appeared in the formula for the angular velocity of precession.

5.1.4 Vertical component of the tidal acceleration

The vertical component of the tidal acceleration causes variations of the gravity acceleration. Denoting this vertical component by Z, and considering it to be positive upward (Fig. 5.5), formulae (5.18) and (5.19) yield

$$Z = \frac{\partial V_t}{\partial r} = \frac{GMr}{2R^3}(3\cos 2\theta + 1)\ . \tag{5.24}$$

The resultant gravity acceleration is then $g - Z$, where g is the gravity acceleration without tides. Using parameters (5.22) and the mean radius of the Earth, $r = 6\,371$ km, we arrive at the following numerical values (in $\mu\mathrm{m\,s^{-2}}$):

$$\begin{aligned} Z_M &= 0.82\left(\cos 2\theta_M + \frac{1}{3}\right)\ , \\ Z_S &= 0.38\left(\cos 2\theta_S + \frac{1}{3}\right)\ . \end{aligned} \tag{5.25}$$

The maximum difference for the lunar tides is $1.65\mu\text{m s}^{-2}$ and for the solar tides $0.76\mu\text{m s}^{-2}$; these values are the doubles of the coefficients in (5.25), as the cosines may attain the values between –1 and +1. If the lunar and solar tides add up, the maximum difference could be $2.41\mu\text{m s}^{-2}$. These differences can easily be detected by gravimeters and are of the same order of magnitude as the gravity anomalies considered in gravimetric prospecting. Consequently, the tidal acceleration must be taken into account in these measurements.

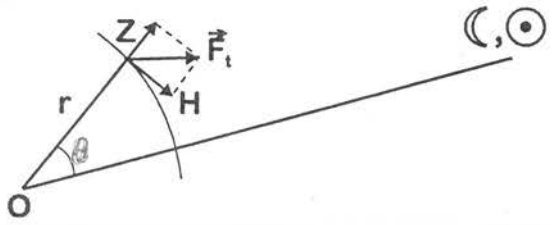

Figure 5.5: Decomposition of the tidal acceleration (F_t) into the vertical (Z) and horizontal (H) components.

5.1.5 Horizontal component of the tidal acceleration

Denote this component by H and consider it to be positive in the direction of decreasing θ, i.e. towards the perturbing body (Fig. 5.5). Then

$$H = -\frac{1}{r}\frac{\partial V_t}{\partial \theta} = \frac{3}{2}\frac{GMr}{R^3}\sin 2\theta \ . \tag{5.26}$$

These horizontal accelerations cause tilts of the vertical. The tilt, $\in$, being a small quantity, can be expressed as

$$\in \approx \tan \in = \frac{H}{g} \ . \tag{5.27}$$

Considering the mean gravity acceleration $g = 9.81\ \text{m s}^{-2}$, we arrive at

$$\begin{aligned} \in_M &= 0.017 \sin 2\theta_M \quad (\text{arc seconds}) \ , \\ \in_S &= 0.008 \sin 2\theta_S \quad (\text{arc seconds}) \ . \end{aligned} \tag{5.28}$$

These tilts represent measurable quantities, so that they must be taken into account in precise astronomical and geodetic measurements.

The instruments used for measuring tilts are called tiltmeters. The classical tiltmeters usually work on the principle of the horizontal pendulum (Zöllner pendulum and others). Many modern instruments measure tilts in terms of

electrical quantities, such as electric current or capacity; for detailed descriptions we refer the reader to Melchior (1966, 1983), and Pick et al. (1973).

5.1.6 Tidal deformations of equipotential surfaces

Consider a small change of the gravity potential, dU, corresponding to a small distance, ds, along the vertical. Then $|\mathrm{d}U| = g|\mathrm{d}s|$, where g is the gravity acceleration.

Consequently, the vertical displacement of an equipotential surface due to tides can be expressed approximately as

$$\zeta = \frac{V_t}{g} \,. \tag{5.29}$$

A more detailed derivation of the latter formula will be given below. If the sign of the tidal potential is given according to (5.18), the places of high tides are characterised by positive values of V_t. Consequently, positive values of ζ describe high tides, and negative ζ characterise low tides.

Substituting (5.20) to (5.22) into (5.29), and taking $r = 6\,371$ km, $g = 9.81\ \mathrm{m\ s^{-2}}$, we arrive at the tidal deformations of the equipotential surfaces near the Earth's surface as follows:

$$\zeta_M = 26.8\left(\cos 2\theta_M + \frac{1}{3}\right) \quad \text{in centimetres} \,,$$

$$\zeta_S = 26.8\left(\cos 2\theta_S + \frac{1}{3}\right) \quad \text{in centimetres} \,. \tag{5.30}$$

Thus, the maximum difference in the height due to the lunar tide is 53.6 cm, and due to the solar tide 24.6 cm. When the maximum tides due to the Moon and Sun coincide, the maximum difference in height is 78.2 cm.

Tides would have the amplitudes given above, if the Earth were ideally fluid, and if there were no dynamic effects (resonances or delays due to inertia, and if the attraction of the displaced masses were neglected as well). If the Earth were rigid, its surface would not deform at all. As the physical properties of the Earth are somewhere between these two extreme cases, the surface is partly deformed. The magnitude of this deformation is determined by the physical properties of the Earth. Therefore, the measurements of the deformations of the Earth's surface contribute to the knowledge of the physical parameters of the Earth. It has been found that the effective deformation of the Earth's surface amounts to about 48 % of the theoretical deformations of the equipotential surfaces mentioned above (theoretical geoidal deformations). Thus the maximum change of the vertical position on the Earth's surface is about 36 cm; further details will be given below.

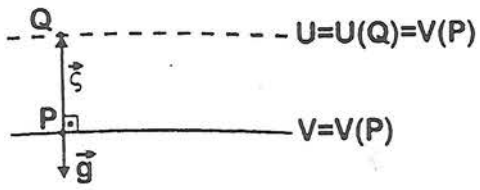

Figure 5.6: Geometry for the derivation of the tidal deformation ζ.

Let us return to a more exact derivation of formula (5.29) for the tidal deformation. Consider an equipotential surface of the gravity field without tides, passing through point P (the solid line in Fig. 5.6). Denote this gravity potential by V. By adding tidal potential V_t (or another disturbing potential), the equipotential surface of the resultant potential, $U = V + V_t$, will shift (dashed line). As the vector of displacement ζ is small,

$$V(Q) - V(P) \doteq +\mathbf{g} \cdot \zeta = -g\zeta \ , \tag{5.31}$$

if g is considered to be positive downward and ζ upward. For the disturbed field similarly

$$U(Q) - U(P) \doteq (\mathbf{g} + \mathbf{F}_t) \cdot \zeta \doteq -g\zeta \ . \tag{5.32}$$

Since $U = V + V_t$ and $U(Q) = V(Q)$, the last equation yields

$$V(P) - [V(P) + V_t(P)] = -g\zeta \ ,$$

from which we immediately obtain equation (5.29).

Equation (5.29) has the same form as Bruns' theorem, which will be derived in the chapter on the geoid and which also relates the deformation of an equipotential surface to a certain type of disturbing potential.

5.2 Angular Distance of Two Points on a Sphere

To study other tidal phenomena, we shall derive a very important formula of spherical trigonometry for the angular distance of two points on a sphere.

We shall consider the general case of the angle between two arbitrary radius-vectors. Consider two arbitrary points, P_1 and P_2, in three-dimensional space, and denote γ the angle between their radius-vectors (Fig. 5.7). Denote the Cartesian coordinates of P_1 and P_2 by x_1, y_1, z_1 and x_2, y_2, z_2, respectively, and their spherical coordinates by r_1, θ_1, λ_1 and r_2, θ_2, λ_2, respectively. The following well-known relations hold true:

$$x_1 = r_1 \sin\theta_1 \cos\lambda_1 \ , \quad x_2 = r_2 \sin\theta_2 \cos\lambda_2 \ ,$$

$$y_1 = r_1 \sin\theta_1 \sin\lambda_1 \ , \quad y_2 = r_2 \sin\theta_2 \sin\lambda_2 \ , \tag{5.33}$$

$$z_1 = r_1 \cos\theta_1 \ , \quad z_2 = r_2 \cos\theta_2 \ .$$

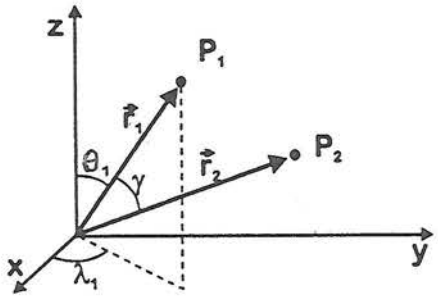

Figure 5.7: Angle γ between two radius-vectors.

The corresponding radius-vectors of points P_1 and P_2 are:

$$\mathbf{r}_1 = (x_1, y_1, z_1) \ , \quad \mathbf{r}_2 = (x_2, y_2, z_2) \ . \tag{5.34}$$

It is evident that $|\mathbf{r}_1| = \sqrt{x_1^2 + y_1^2 + z_1^2} = r_1$ and $|\mathbf{r}_2| = r_2$. We would like to express the angle γ between these radius-vectors in terms of the spherical coordinates of points P_1 and P_2.

Calculate the scalar product of vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, and substitute from (5.33):

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = x_1 x_2 + y_1 y_2 + z_1 z_2 = r_1 r_2 \left[\cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\lambda_1 - \lambda_2)\right] . \tag{5.35}$$

However, from analytical geometry we know that

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = |\mathbf{r}_1||\mathbf{r}_2| \cos\gamma \ . \tag{5.36}$$

By comparing (5.35) and (5.36) we arrive at the final result:

$$\boxed{\cos\gamma = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\lambda_1 - \lambda_2)} \ . \tag{5.37}$$

5.3 Three Types of Tides

A very restrictive assumption in Section 5.1 was that the Earth does not rotate about its axis. Now, let us abandon this restriction and assume that a rigid Earth also rotates about its axis. To the forces acting at point P we must add the

centrifugal force due to the Earth's rotation. However, this force field is included in the gravity field of the rotating Earth. This means that the values of g, considered above, must be slightly modified by this centrifugal acceleration. Consequently, the only substantial change consists in angle θ (see Fig. 5.2 and the formulae in Section 5.1), becoming a very complicated function of time.

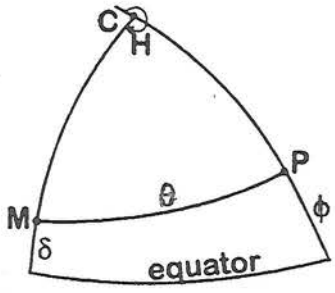

Figure 5.8: Definition of quantities which appear in the formulae for the tides on a rotating Earth.

Consider the tides due to the Moon. If we permit the Earth to rotate, an observer at P moves through a varying potential field, and the potentials become functions of time. In Fig. 5.8, point C is the pole of the celestial sphere, M is the Moon, and P is the point on the Earth. The position of the Moon on the celestial sphere with respect to the observer can be characterised by the declination, δ, and by the Moon's hour angle, H. Denote the geocentric latitude of P by ϕ. Then, by applying formula (5.37) to spherical triangle CMP, the Moon's geocentric angular distance θ can be expressed as

$$\cos\theta = \sin\phi\sin\delta + \cos\phi\cos\delta\cos(H - 180°) \ . \tag{5.38}$$

After substituting this expression into tidal potential (5.14), changing the sign of the potential and using relation $\cos^2 H = (1+\cos 2H)/2$, we arrive at the leading term of the tidal potential in the form

$$V_t = \frac{3}{4}\frac{GMr^2}{R^3}\Big[\cos^2\phi\cos^2\delta\cos 2H - \\ - \sin 2\phi\sin 2\delta\cos H + 3\left(\sin^2\phi - \frac{1}{3}\right)\left(\sin^2\delta - \frac{1}{3}\right)\Big] \ . \tag{5.39}$$

The three terms on the right-hand side of (5.39) represent three types of tides on the rotating Earth due to the Moon. A similar expression can be given for the tidal potential due to the Sun.

The individual terms in (5.39) vary differently with time, and with the latitude of P. The first term, containing $\cos 2H$, yields the semi-diurnal tide (the period of half a lunar or solar day). This term is symmetric about the equator.

In the second term, factor $\cos H$ indicates a period of one (lunar or solar) day, corresponding to the diurnal tide. This term is antisymmetric about the equator, and vanishes if the disturbing body or the observer are at the equator. At first sight, the appearance of such a diurnal term seems to be surprising, but this can easily be explained (Fig. 5.9). We have emphasized the symmetry of tides with respect to plane $\theta = 90°$, which means that the tides at points P and Q are the same. If the disturbing body lies in the equatorial plane of the Earth (Fig. 5.9a), the observer at point P appears half a day later at point Q with the same tidal potential. This situation is described by the semi-diurnal term. However, if the rotation axis is inclined (Fig. 5.9b), the observer at P then appears half a day later at point P^*, where the tidal potential is different. Only one day later does the observer return to the place of the same potential at P. This explains the diurnal tide.

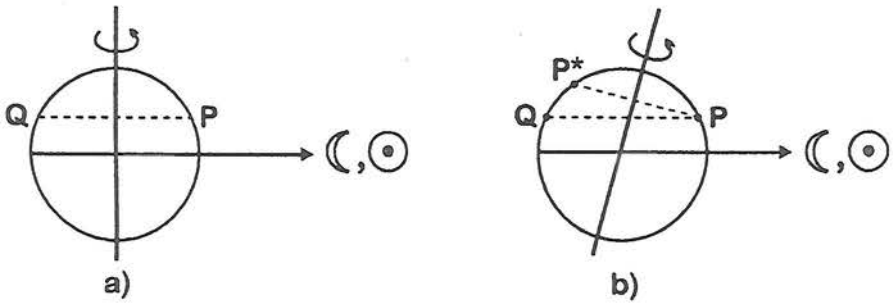

Figure 5.9: Explanation of the diurnal component of tides.

The third term in (5.39), varying only with δ, yields tides of a long period. This term has a period of 14 days in the lunar tides, and a period of 6 months in the solar tides. A remarkable property of this term is that it yields a small permanent high tide in the equatorial regions, and a permanent low tide in the polar regions. Consequently, it slightly increases the flattening of the Earth. The equipotential surface at the poles is displaced downwards by 28 cm, and elevated by 14 cm at the equator. These values are small, but they play a certain role in accurate determinations of the Earth's flattening and in some other parameters. Therefore, it must be specified whether the effect of this term is included or not. For example, the polar flattening of the Earth including this term is $\alpha = 1/(298.256\,42 \pm 0.000\,01)$, but without this term it is (the tide-free value) $\alpha = 1/(298.257\,65 \pm 0.000\,01)$. The permanent polar depression due to this term can again be easily explained. We have seen that the high tide occurs for angular distance $\theta < 54.7°$ (and symmetrically on the opposite side of the Earth). Even if the Sun is maximally deviated from the Earth's equatorial plane by $\epsilon = 23.4°$, the high tide reaches only to a latitude of $54.7° + 23.4° = 78.1°$.

Consequently, a certain vicinity of the poles remains in the region of the low tide all the time. In addition to these effects, the long-period variations due to this term cause changes in the principal moment of inertia C. Consequently, we should expect variations in the speed of the Earth's rotation to have the periods mentioned above. Such variations in the length of day have really been observed.

Figure 5.10 shows the signs of the individual terms in (5.39).

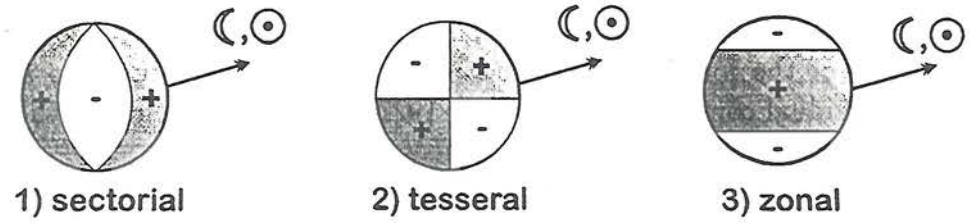

Figure 5.10: Three types of tides.

In reality the tides are even more complicated due the ellipticities of the orbits and perturbations of the Moon's motion. Consequently, the harmonic analysis of the forces and *harmonic analysis of tides* are used to separate the individual components (Doodson, 1921; Garland, 1965; Pick et al., 1973; Melchior, 1983).

5.4 Tidal Effects on an Elastic Earth

5.4.1 Love numbers

If an elastic Earth is considered, the tidal forces cause not only deformations of the equipotential surfaces, but also deformations of the Earth's surface.

Denote by U the gravity potential, and by g the gravity acceleration if there are no tides. Denote by V_t the tidal potential for a rigid Earth (tide-generating potential), which has been considered in the previous sections. The resultant gravity potential, including the tidal contributions, at a point of an elastic Earth can then be expressed as

$$U_{el} = U + V_t + V_t' - V_t'' \,, \tag{5.40}$$

where V_t' is the additional potential at the point due to the gravitational effect of the transferred mass, and V_t'' is the change in the potential due to the vertical displacement u of the observer.

Love (1909) showed that there is a proportionality between the additional and generating potentials,

$$V_t' = kV_t \,, \tag{5.41}$$

and between the vertical deformation of the Earth's surface and the theoretical vertical deformation of the equipotential surface, ζ, mentioned above,

$$u = h\zeta \ . \tag{5.42}$$

In other words, number h is the ratio of the height of the Earth's tides to the height of the corresponding static oceanic tides.

Constants h and k are called the *Love numbers*. In addition to them, also the so-called *Shida number*, l, is introduced as the ratio of the horizontal displacement in the solid Earth to the horizontal displacement of the corresponding static oceanic tides. The Love and Shida numbers characterise the deformations of the Earth due to the tidal forces, i.e. they characterise *elastic properties of the Earth* as a whole. They depend on the density distribution and elastic parameters within the Earth. More detailed calculations show that these numbers also slightly depend on the period of the tidal component.

5.4.2 Determination of the Love numbers

Using (5.42) and (5.29), potential V_t'' can be expressed as

$$V_t'' = gu = gh\zeta = hV_t \ . \tag{5.43}$$

The resultant gravity potential (5.40) then becomes

$$U_{el} = U + DV_t \ , \tag{5.44}$$

where

$$\boxed{D = 1 + k - h} \ . \tag{5.45}$$

Parameter D is called the diminishing factor, because $D < 1$ as indicated by measurements. Diminishing factor D can be determined by measuring the tilts with tiltmetres. The measurements yield

$$D \doteq 0.7 \ . \tag{5.46}$$

Another combination of the Love numbers can be determined by gravity measurements. The change in gravity on an elastic Earth can be expressed as

$$\delta g = G^* \frac{\partial V_t}{\partial r} \ , \tag{5.47}$$

where gravimetric factor G^* (magnification factor) is

$$\boxed{G^* = 1 + h - \frac{3}{2}k}\,; \tag{5.48}$$

we shall not derive the latter formula. (Note that the gravimetric factor is usually denoted by G, but we have reserved the letter G for the gravitational constant). This factor can be determined by means of sensitive gravimeters. Its approximate value is

$$G^* \doteq 1.2\ . \tag{5.49}$$

Equations (5.46) and (5.49) then yield

$$h \doteq 0.5, \quad k \doteq 0.2\ . \tag{5.50}$$

However, there is a large scatter in the values of D and G^* determined at various observatories.

According to Melchior (1966), the following numerical values seem to be more probable:

$$\begin{aligned} &D = 0.68\ ,\ G^* = 1.17\ , \\ &h = 0.62\ ,\ k = 0.30\ , \\ &l = 0.05\ . \end{aligned} \tag{5.51}$$

Another set of these numbers is as follows (Wahr, 1996):

$$\begin{aligned} &h \approx 0.6\ , \quad k \approx 0.3 \\ &l \approx 0.085\ ,\ G^* = 1.16 \text{ (also denoted as } \delta = 1.16)\ . \end{aligned}$$

Love number k can also be determined independently from the Chandler period, from the variations of the Earth's rotation, and from the tidal perturbation of satellite orbits. The last method is the most accurate, yielding the value $k = 0.290$.

The Love numbers are related to the density distribution and elastic parameters within the Earth. Hence, these numbers can be calculated for different models of the Earth, and compared with observations. The first calculations were carried out by Takeuchi in 1950. It was found that an estimate of rigidity in the liquid core can be obtained by this method. Takeuchi concluded that the rigidity in the core must be at least three orders lower than that in the mantle ($\mu \sim 0 - 10^8$ N m^{-2} in the core satisfied the data; $\mu \sim 10^{11}$ N m^{-2} in the mantle).

In conclusion we should summarize the effects which were not considered in this chapter. These are: dynamic effects in the ocean, dynamic effects in the solid Earth (we have considered static tides only, which is substantiated by the fact that the tidal periods are longer than the periods of the free oscillations of the Earth), local effects and some others.

Chapter 6

Legendre Polynomials and Associated Legendre Functions

In many branches of physics, we encounter potentials which vary as $1/R$, where R is the distance from the source. In many applications, it is convenient to expand such an expression into a series. The coefficients of such expansions contain special functions called Legendre polynomials. In the previous chapter, we have already seen the advantages of such an approach. Only one term of the series was sufficient to describe the main properties of tides.

We shall start with a very special situation, which will be generalised later in this chapter.

Legendre polynomials may be introduced in several equivalent ways: as particular solutions of Legendre's differential equation, by means of recurrence relations, integral expressions, Rodrigues' formula, generating function, in the form of a finite series and some others. If any of these representations is adopted as a definition of Legendre polynomials, the remaining representations can be derived from this definition. Here we shall introduce Legendre polynomials by means of a generating function, because such an approach has a very simple physical interpretation.

6.1 Generating Function

Consider a special position of a point source of the gravitational or electrostatic field, B, on the z-axis of a Cartesian coordinate system at unit distance from the origin, i.e. $B = B(0, 0, 1)$. We can imagine that this point is located at the "north" pole of a unit sphere with its centre at the origin of the coordinate system (Fig. 6.1).

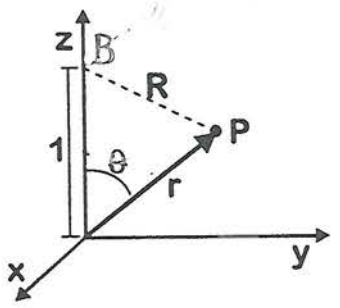

Figure 6.1: Introduction of the generating function of Legendre polynomials.

Consider another point, P (observer), inside the unit sphere, and denote its spherical coordinates by r, θ, λ, where $r<1$. Distance R between points P and B, using the law of cosines (cosine rule), satisfies

$$R^2 = 1 + r^2 - 2r\cos\theta \ , \tag{6.1}$$

and, consequently, the reciprocal distance can be expressed as

$$\frac{1}{R} = \frac{1}{\sqrt{1-2r\cos\theta + r^2}} \ . \tag{6.2}$$

Put $\mu = \cos\theta$. Now expand the reciprocal distance into a series in powers of r,

$$\boxed{\frac{1}{\sqrt{1-2r\mu + r^2}} = \sum_{n=0}^{\infty} r^n P_n(\mu)} \ , \tag{6.3}$$

the coefficient of the n-th power being denoted $P_n(\mu)$. We shall show that $P_n(\mu)$ are polynomials, so-called Legendre polynomials. The left-hand side of equation (6.3) is called the *generating function* of Legendre polynomials:

$$\boxed{g(r,\mu) = \frac{1}{\sqrt{1-2r\mu + r^2}}} \ . \tag{6.4}$$

We consider equation (6.3) as the definition of Legendre polynomials in terms of the generating function.

Let us derive the first few coefficients, Legendre polynomials, of the corresponding series. Without expressing variable μ explicitly for the moment, the appropriate Taylor series reads

$$g(r) = g(0) + \frac{g'(0)}{1!}r + \frac{g''(0)}{2!}r^2 + \dots \ , \tag{6.5}$$

where

$$g(0) = 1 \ ,$$

$$g'(r) = -\frac{1}{2}\left(1-2r\mu + r^2\right)^{-\frac{3}{2}}\left(-2\mu + 2r\right), \quad g'(0) = \mu \ ,$$

$$g''(r) = -\frac{1}{2}\left(-\frac{3}{2}\right)\left(1-2r\mu + r^2\right)^{-\frac{5}{2}}\left(-2\mu + 2r\right)^2 -$$

$$-\left(1-2r\mu+r^2\right)^{-\frac{3}{2}}, \quad g''(0)=3\mu^2-1 .$$

Comparison of (6.3) with (6.5) yields

$$\boxed{P_n(\mu)=\frac{1}{n!}\left[\frac{\partial^n g}{\partial r^n}\right]_{r=0}} \tag{6.6}$$

and, consequently,

$$P_0(\mu)=1 ,$$

$$P_1(\mu)=\mu , \tag{6.7}$$

$$P_2(\mu)=\frac{3}{2}\mu^2-\frac{1}{2} .$$

Continuing, we would obtain

$$P_3(\mu)=\frac{5}{2}\mu^3-\frac{3}{2}\mu$$

and so on; see the graphs in Fig. 6.2.

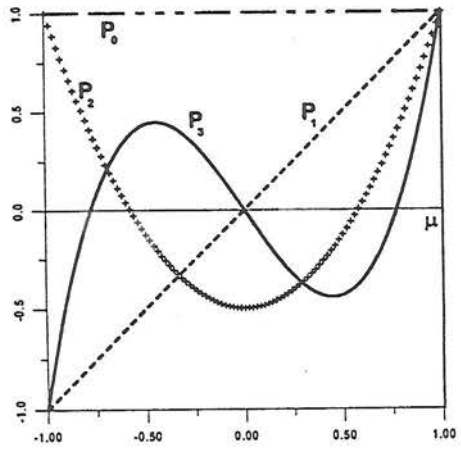

Figure 6.2: Legendre polynomials $P_n(\mu)$.

Let us add two remarks:

1. Although the polynomials on the right-hand sides of (6.7) are defined along the whole real axis, Legendre polynomials are defined only in the

interval $\langle -1, +1 \rangle$ as a consequence of the geometric meaning of variable μ, i.e. $\mu = \cos\theta$.

2. Functions P_n as functions of variable θ are not polynomials, but more complicated functions (Fig. 6.3).

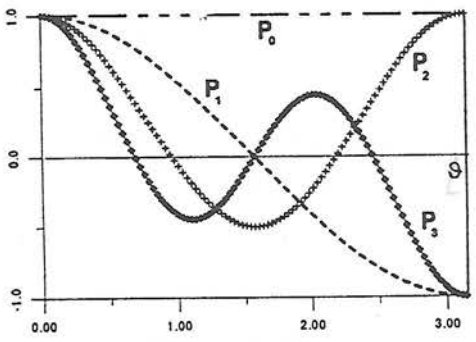

Figure 6.3: Functions $P_n(\cos\theta)$.

6.2 Some Values and Special Properties

The graphs in Fig. 6.2 indicate some of the general properties of Legendre polynomials, which will now be proved using definition (6.3):

a) Values for $\mu = 1$. Inserting $\mu = 1$ into equation (6.3) yields

$$\frac{1}{1-r} = \sum_{n=0}^{\infty} r^n P_n(1) .$$

However, the left-hand side of the latter equation is the sum of the geometric series,

$$\frac{1}{1-r} = \sum_{n=0}^{\infty} r^n .$$

By comparing the right-hand sides, we see that

$$\boxed{P_n(1) = 1} \qquad (6.8)$$

for all values of n.

b) Values for $\mu = -1$. Definition (6.3) yields

$$\frac{1}{1+r} = \sum_{n=0}^{\infty} r^n P_n(-1) .$$

The left-hand side is also the sum of a geometric series,

$$\frac{1}{1-(-r)} = \sum_{n=0}^{\infty}(-r)^n \ .$$

Hence

$$\boxed{P_n(-1) = (-1)^n} \ . \tag{6.9}$$

c) Parity. Replace μ by $(-\mu)$ in equation (6.3):

$$\frac{1}{\sqrt{1+2r\mu+r^2}} = \sum_{n=0}^{\infty} r^n P_n(-\mu) \ .$$

But the left-hand side may be expressed as

$$\frac{1}{\sqrt{1-2(-r)\mu+(-r)^2}} = \sum_{n=0}^{\infty}(-r)^n P_n(\mu) \ .$$

Consequently,

$$\boxed{P_n(-\mu) = (-1)^n P_n(\mu)} \ . \tag{6.10}$$

Thus, Legendre polynomials are even functions (with respect to $\mu = 0$) for even degrees n, and odd functions for odd degrees.

6.3 Recurrence Relations

Differentiate the basic equation (6.3) with respect to r,

$$\frac{\mu - r}{\left(1-2\mu r+r^2\right)^{3/2}} = \sum_{n=1}^{\infty} n r^{n-1} P_n(\mu) \ ,$$

and multiply both sides by $\left(1-2r\mu+r^2\right)$. By substituting (6.3) into the left-hand side we arrive at

$$(\mu - r)\sum_{n=0}^{\infty} r^n P_n(\mu) = \left(1-2r\mu+r^2\right)\sum_{n=1}^{\infty} n r^{n-1} P_n(\mu) \ .$$

Comparison of the coefficients of terms r^n yields

$$\mu P_n(\mu) - P_{n-1}(\mu) = (n+1)P_{n+1}(\mu) - 2n\mu P_n(\mu) + (n-1)P_{n-1}(\mu) \ .$$

Consequently, we arrive at the recurrence relation

$$\boxed{(n+1)P_{n+1}(\mu) - (2n+1)\mu P_n(\mu) + nP_{n-1}(\mu) = 0} \ . \qquad (6.11)$$

If P_{n-1} and P_n are known, this relation can easily be used to calculate P_{n+1}. Using $P_0(\mu) = 1$ and $P_1(\mu) = \mu$, apply this formula to obtain $P_2(\mu)$, $P_3(\mu)$, etc.

Relation (6.11) also indicates that all $P_n(\mu)$ are polynomials. Namely, assuming $P_{n-1}(\mu)$ to be a polynomial of the (n–1)-th degree and $P_n(\mu)$ to be a polynomial of the n-th degree, it follows from relation (6.11) that $P_{n+1}(\mu)$ is a polynomial of degree (n+1).

Another relation can be obtained if we differentiate equation (6.3) with respect to μ. This gives

$$\frac{r}{\left(1-2r\mu+r^2\right)^{3/2}} = \sum_{n=0}^{\infty} r^n P_n'(\mu) \ ,$$

or after multiplying by $\left(1-2r\mu+r^2\right)$,

$$r\sum_{n=0}^{\infty} r^n P_n(\mu) = \left(1-2r\mu+r^2\right)\sum_{n=0}^{\infty} r^n P_n'(\mu) \ .$$

Comparison of the coefficients of terms r^{n+1} now yields

$$P_{n+1}'(\mu) + P_{n-1}'(\mu) = 2\mu P_n'(\mu) + P_n(\mu) \ .$$

Numerous additional equations may be developed from this equation, equation (6.11) and its derivatives (Arfken, 1970).

6.4 Legendre's Differential Equation

Put

$$z = \sqrt{1-2r\mu+r^2} \ ,$$

$$g = \frac{1}{z} = \frac{1}{\sqrt{1-2r\mu+r^2}} \ .$$

The derivatives of these functions read

$$\frac{\partial z}{\partial \mu} = \frac{1}{z}(-r) \ , \quad \frac{\partial z}{\partial r} = \frac{1}{z}(r - \mu) \ ,$$

$$\frac{\partial^2 z}{\partial \mu^2} = -\frac{1}{z^2}\left(\frac{-r}{z}\right)(-r) = -r^2 g^3 \ ,$$

$$\frac{\partial^2 z}{\partial r^2} = -\frac{1}{z^2}\frac{(r-\mu)^2}{z} + \frac{1}{z} = \frac{1-\mu^2}{z^3} = (1-\mu^2)g^3 \ .$$

Consider the last two equations. To obtain the same expressions on their right-hand sides, multiply the first equation by $(1-\mu^2)$, and the second by r^2. By adding them we get

$$(1-\mu^2)\frac{\partial^2 z}{\partial \mu^2} + r^2 \frac{\partial^2 z}{\partial r^2} = 0$$

Differentiate this with respect to μ and substitute $\frac{\partial z}{\partial \mu} = -rg$:

$$(1-\mu^2)(-r)\frac{\partial^2 g}{\partial \mu^2} - 2\mu(-r)\frac{\partial g}{\partial \mu} + r^2 \frac{\partial^2}{\partial r^2}(-rg) = 0 \ .$$

Dividing by $(-r)$ yields

$$(1-\mu^2)\frac{\partial^2 g}{\partial \mu^2} - 2\mu\frac{\partial g}{\partial \mu} + r\frac{\partial^2}{\partial r^2}(rg) = 0 \ .$$

By substituting

$$g = \sum_{n=0}^{\infty} r^n P_n(\mu) \tag{6.12}$$

and comparing the coefficients, we arrive at *Legendre's differential equation* for $P_n(\mu)$:

$$(1-\mu^2)\frac{\mathrm{d}^2 P_n(\mu)}{\mathrm{d}\mu} - 2\mu\frac{\mathrm{d} P_n(\mu)}{\mathrm{d}\mu} + n(n+1)P_n(\mu) = 0 \ , \tag{6.13a}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left[(1-\mu^2)\frac{\mathrm{d} P_n(\mu)}{\mathrm{d}\mu}\right] + n(n+1)P_n(\mu) = 0 \ . \tag{6.13b}$$

6.5 Bounds for Legendre Polynomials

Consider the binomial expansion

$$(1+x)^m = 1+mx+\frac{m(m-1)}{2!}x^2+\frac{m(m-1)(m-2)}{3!}x^3+\ldots \qquad (6.14)$$

which is convergent for x satisfying the relation $-1<x<1$. Using (6.14) we can expand the generating function for Legendre polynomials as follows:

$$\left(1-2r\cos\theta+r^2\right)^{-1/2} = \left(1-re^{i\theta}\right)^{-1/2}\left(1-re^{-i\theta}\right)^{-1/2} = \qquad (6.15)$$

$$= \left(1+\frac{1}{2}re^{i\theta}+\frac{3}{8}r^2e^{2i\theta}+\ldots\right)\left(1+\frac{1}{2}re^{-i\theta}+\frac{3}{8}r^2e^{-2i\theta}+\ldots\right),$$

all coefficients being positive and decreasing. Both series are absolutely convergent for $|r|<1$ as

$$1+\left|\frac{1}{2}re^{\pm i\theta}\right|+\left|\frac{3}{8}r^2e^{\pm 2i\theta}\right|+\ldots \le 1+r+r^2+\ldots = \frac{1}{1-r}\,.$$

Since the absolutely convergent series may be multiplied term by term, equation (6.15) yields

$$\left(1-2r\cos\theta+r^2\right)^{-1/2} = 1+\frac{1}{2}r\left(e^{i\theta}+e^{-i\theta}\right)+\frac{1}{4}r^2e^{i\theta}e^{-i\theta}+\frac{3}{8}r^2\left(e^{2i\theta}+e^{-2i\theta}\right)+\ldots \qquad (6.16)$$

$$= 1+r\cos\theta+r^2\left(\frac{1}{4}\cos 0+\frac{3}{4}\cos 2\theta\right)+\ldots\,.$$

The Legendre polynomial is the coefficient of r^n, i.e.

$$P_n(\cos\theta) = \sum_{m=0 \text{ or } 1}^{n} a_m \cos m\theta\,, \qquad (6.17)$$

all the a_m being positive. This series has a maximum if $\cos m\theta = 1$, i.e. $\theta = 0$. Therefore,

$$\boxed{\left|P_n(\cos\theta)\right| \le P_n(1) = 1}\,, \qquad (6.18)$$

where we have used equation (6.8).

The latter relation is very important in estimating the convergence of series containing Legendre polynomials. We could have already expected this relation to hold true when drawing the graphs of Legendre polynomials in Section 6.1. Now, we have proved it. Note that equation (6.17) could be another formula for calculating Legendre polynomials if suitable algebraic formulae for coefficients a_m are available.

6.6 Other Properties of Legendre Polynomials

In this section, we shall mention several other important formulae for Legendre polynomials. However, since we shall not need them in subsequent chapters, we shall give them without proofs.

The simplest formula for Legendre polynomials to remember is Rodrigues' formula,

$$P_n(\mu) = \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}\mu^n}\left[\left(\mu^2 - 1\right)^n\right] . \tag{6.19}$$

The straightforward derivation of this formula is based on the theory of functions of a complex variable. Starting with formula (6.6), the derivative in this formula is expressed in the form of Cauchy's integral, which is then calculated by means of the residue theorem. It immediately yields formula (6.19). A more elementary, but longer derivation of Rodrigues' formula can be found, e.g., in (Arfken, 1970).

Legendre polynomials are orthogonal for the interval $\langle -1, 1 \rangle$, i.e.

$$\int_{-1}^{1} P_n(\mu) P_m(\mu) \,\mathrm{d}\mu = 0 \tag{6.20}$$

if $n \neq m$. If $n = m$,

$$\int_{-1}^{1} \left[P_n(\mu)\right]^2 \,\mathrm{d}\mu = \frac{2}{2n+1} . \tag{6.21}$$

Elegant proofs of these formulae can also be found in (Arfken, 1970).

6.7 Expansion of the Reciprocal Distance of Two Arbitrary Points

In Section 6.1 we considered the distance (and also the reciprocal distance) of two points, one of which was located at the "north pole" of a unit sphere, and the second was inside the sphere.

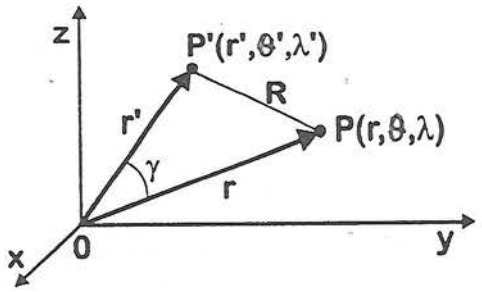

Figure 6.4: Coordinates of two arbitrary points and their distance.

Now, let us consider a general case of two arbitrary points, P and P' (Fig. 6.4). Denote the spherical coordinates of point P by r, θ, λ, and of point P' by r', θ', λ'. The distance between the points, according to the law of cosines, is

$$R = \sqrt{r^2 + r'^2 - 2rr'\cos\gamma} \ , \tag{6.22}$$

where γ is the angle between the radius-vectors of the points. If $r > r'$,

$$\frac{1}{R} = \frac{1}{r}\frac{1}{\sqrt{1 - 2\frac{r'}{r}\cos\gamma + \left(\frac{r'}{r}\right)^2}} \ . \tag{6.23a}$$

If $r < r'$, we shall exchange r and r':

$$\frac{1}{R} = \frac{1}{r'}\frac{1}{\sqrt{1 - 2\frac{r}{r'}\cos\gamma + \left(\frac{r}{r'}\right)^2}} \ . \tag{6.23b}$$

By comparing these expressions with the generating function, see equations (6.2) and (6.3), we arrive at

$$\boxed{\begin{aligned} \frac{1}{R} &= \frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n P_n(\cos\gamma) \qquad \text{for } r > r' \ , \\ \frac{1}{R} &= \frac{1}{r'}\sum_{n=0}^{\infty}\left(\frac{r}{r'}\right)^n P_n(\cos\gamma) \qquad \text{for } r < r' \ . \end{aligned}} \tag{6.24a, 6.24b}$$

These expressions belong to the *most important* expressions in the theory of Legendre polynomials. They represent the required generalisation of the problem

which was discussed in Section 6.1. Formulae (6.2) and (6.3) of Section 6.1 are special cases of (6.24b) for $r' = 1$ and $\theta' = 0$.

Note that the larger of the coordinates r and r' always appears in the denominators of formulae (6.24). Together with the limitation (6.18) of Legendre polynomials, i.e. $\left|P_n(\cos\gamma)\right| \le 1$, it ensures the convergence of the corresponding series.

Formulae (6.24) do not describe the case of $r = r'$. To tell the truth, we shall try to avoid this case, if possible, since problems appear with the convergence of the corresponding series. For example, the series is divergent for $\gamma = 0$ and $\gamma = \pi$ (oscillatory), which immediately follows from (6.8) and (6.9).

Expansions (6.24) allow reciprocal distance $1/R$ to be expressed in terms of distances r and r', and of the angle γ between the radius-vectors. Such an expansion is fully sufficient in many applications. For example, we used it (only with a different notation) in the derivation of the tidal potential for a non-rotating Earth. However, in many other applications we need reciprocal distance $1/R$ to be expressed in terms of the spherical coordinates of points P and P'. In principle, this can easily be done if we substitute $\cos\gamma$ with the trigonometric identity

$$\boxed{\cos\gamma = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\lambda - \lambda')}\,; \qquad (6.25)$$

see the derivation of this formula in Chapter 5. We have used this approach in deriving the tidal potential for a rotating Earth. However, this simple method leads to very complicated expressions when the Legendre polynomials of a higher degree n, i.e. higher powers of $\cos\gamma$, are required. A more convenient approach uses the associated Legendre functions and the addition theorem for Legendre polynomials. We shall, therefore, deal with them in the next sections.

6.8 Associated Legendre Functions

The associated Legendre function of degree n and order m is defined by

$$\boxed{P_n^m(\mu) = \left(1-\mu^2\right)^{m/2} \frac{\mathrm{d}^m P_n(\mu)}{\mathrm{d}\mu^m}}\,, \qquad (6.26)$$

where $P_n(\mu)$ is the Legendre polynomial. In order to complete this system of functions, also the zero-order associated Legendre functions are defined by

$$P_n^0(\mu) = P_n(\mu)\,, \qquad (6.27)$$

i.e. the zero-th derivative indicates that no differentiation is carried out. Since Legendre polynomial $P_n(\mu)$ is a polynomial of the n-th degree, its m-th derivative will be equal to zero for $m > n$. Therefore,

$$P_n^m(\mu) = 0 \qquad \text{for} \quad m > n \ , \tag{6.28}$$

and we can restrict ourselves only to orders $m = 0, 1, 2, \ldots, n$.

Let us derive expressions for the associated Legendre functions of a few low degrees. In addition to variable μ, we also give the expression in variable θ $(\mu = \cos\theta)$:

$$\begin{aligned}
P_0^0(\mu) &= P_0(\mu) = 1 \ , \\
P_1^0(\mu) &= P_1(\mu) = \mu = \cos\theta \ , \\
P_1^1(\mu) &= \left(1-\mu^2\right)^{1/2} = \sin\theta \ , \\
P_2^0(\mu) &= P_2(\mu) = \frac{3}{2}\mu^2 - \frac{1}{2} = \frac{3}{2}\cos^2\theta - \frac{1}{2} \ , \\
P_2^1(\mu) &= \left(1-\mu^2\right)^{1/2} 3\mu = 3\sin\theta\cos\theta \ , \\
P_2^2(\mu) &= \left(1-\mu^2\right)\cdot 3 = 3\sin^2\theta \ .
\end{aligned} \tag{6.29}$$

It can be seen from these expressions that generally $P_n^m(\mu)$ are not polynomials; they are polynomials only for even m.

6.9 The Addition Theorem for Legendre Polynomials

The addition theorem for Legendre polynomials states that

$$\boxed{P_n(\cos\gamma) = \sum_{m=0}^{n} \epsilon_m \frac{(n-m)!}{(n+m)!} P_n^m(\cos\theta) P_n^m(\cos\theta') \cos\left[m(\lambda-\lambda')\right]} \ , \tag{6.30}$$

where $\epsilon_0 = 1$, $\epsilon_1 = \epsilon_2 = \ldots = 2$, and the notations of Section 6.7 being used.

Let us verify the validity of the addition theorem for the lowest values of n. First, we shall calculate the left-hand side (L.H.S.) of (6.30) by substituting the trigonometric identity (6.25) into expressions (6.7) for Legendre polynomials. We shall then calculate the right-hand side (R.H.S.) of (6.30). For

$n = 0$:

$$\text{L.H.S.} = P_0(\cos\gamma) = 1$$

$$\text{R.H.S.} = P_0^0(\cos\theta) P_0^0(\cos\theta') \cos 0 = 1$$

$n = 1$:

$$\text{L.H.S.} = P_1(\cos\gamma) = \cos\gamma = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\lambda - \lambda')$$

$$\text{R.H.S.} = P_1^0(\cos\theta) P_1^0(\cos\theta') + P_1^1(\cos\theta) P_1^1(\cos\theta')\cos(\lambda - \lambda') =$$
$$= \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\lambda - \lambda')$$

$n = 2$:

$$\text{L.H.S.} = P_2(\cos\gamma) = \frac{3}{2}\cos^2\gamma - \frac{1}{2} =$$

$$= \frac{3}{2}\left[\cos^2\theta\cos^2\theta' + 2\cos\theta\cos\theta'\sin\theta\sin\theta'\cos(\lambda - \lambda') + \right.$$
$$\left. + \sin^2\theta\sin^2\theta'\frac{1 + \cos 2(\lambda - \lambda')}{2}\right] - \frac{1}{2} =$$

$$= \frac{3}{2}\cos^2\theta\cos^2\theta' + \frac{3}{4}\left(1 - \cos^2\theta\right)\left(1 - \cos^2\theta'\right) - \frac{1}{2} +$$
$$+ 3\cos\theta\cos\theta'\sin\theta\sin\theta'\cos(\lambda - \lambda') + \frac{3}{4}\sin^2\theta\sin^2\theta'\cos 2(\lambda - \lambda') =$$

$$= \left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)\left(\frac{3}{2}\cos^2\theta' - \frac{1}{2}\right) +$$
$$+ 3\cos\theta\cos\theta'\sin\theta\sin\theta'\cos(\lambda - \lambda') + \frac{3}{4}\sin^2\theta\sin^2\theta'\cos 2(\lambda - \lambda')$$

$$\text{R.H.S.} = P_2^0(\cos\theta) P_2^0(\cos\theta') + \frac{1}{3}P_2^1(\cos\theta) P_2^1(\cos\theta')\cos(\lambda - \lambda') +$$
$$+ \frac{1}{12}P_2^2(\cos\theta) P_2^2(\cos\theta')\cos 2(\lambda - \lambda') = P_2(\cos\gamma)\ .$$

The complicated algebra, required in the last case, indicates that the addition theorem is a very sophisticated mathematical theorem. Therefore, we cannot expect its general proof to be easy. For this reason, we shall not give this proof here. Actually, in many applications in the following chapters, we shall need this theorem only for $n \leq 2$, and for these values it has just been proved.

Since the addition theorem has been proved for small values of n, we could try to use recurrent formulae for its general proof; see Novotný (1982). For this purpose it would be necessary to derive recurrent formulae also for the associated Legendre functions, and to express the addition theorem (6.30) in a more symmetrical form (to introduce exponential functions instead of the cosine function).

Note that the functions of the type $P_n^m(\cos\theta)\cos m\lambda$ and $P_n^m(\cos\theta)\sin m\lambda$ are called *spherical functions*. Thus, the addition theorem expands $P_n(\cos\gamma)$ in terms of spherical functions (spherical harmonics). The reciprocal distance (6.24a) may now be expressed as

$$\frac{1}{R}=\frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n P_n(\cos\gamma)= \tag{6.31}$$

$$=\frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n\sum_{m=0}^{n}\in_m\frac{(n-m)!}{(n+m)!}P_n^m(\cos\theta)P_n^m(\cos\theta')\cos\left[m(\lambda-\lambda')\right],$$

where $\in_0=1$, $\in_1=\in_2=\ldots=2$. Formula (6.24b) can be modified in a similar way.

Chapter 7

Foundations of the Theory of the Earth's Gravity Field

7.1 Basic Notions

Consider a particle, its mass being denoted by m. According to Newton's law of universal gravitation, the intensity of the gravitational field due to this particle at a distance r is of magnitude

$$E = \frac{Gm}{r^2} \ . \tag{7.1}$$

Introducing the radius vector, **r**, which originates at the particle, the vector of intensity may be expressed as

$$\mathbf{E}(\mathbf{r}) = -\frac{Gm}{r^3}\mathbf{r} \ . \tag{7.2}$$

Its radial component, being considered positive away from the particle,

$$E_r = -\frac{Gm}{r^2} \ . \tag{7.3}$$

Let us introduce the gravitational potential, W, so that

$$\mathbf{E} = -\operatorname{grad} W \ . \tag{7.4}$$

The radial component of intensity is then

$$E_r = -\frac{\partial W}{\partial r} \ , \tag{7.5}$$

and for the gravitational potential of a particle we have

$$W = -\frac{Gm}{r} \ . \tag{7.6}$$

However, in mathematics, astronomy, geodesy and geophysics, we often give the potential the opposite sign, i.e. we change the signs in formulae (7.4) to (7.6). The new function, $U = -W$, is called the potential function, or simply the

potential. Consequently, the *gravitational potential of a particle*, according to the new sign convention, reads

$$U = +\frac{Gm}{r} , \tag{7.7}$$

and the corresponding *intensity of the gravitational field* is

$$\mathbf{E} = +\mathrm{grad}\, U . \tag{7.8}$$

In this chapter we shall use this sign convention.

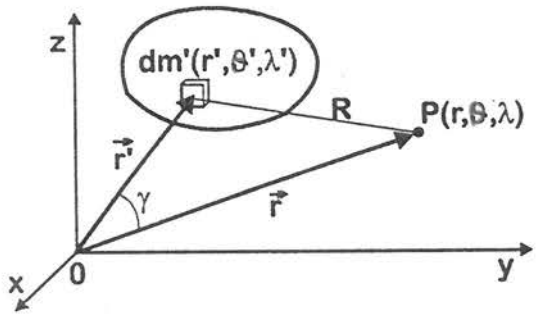

Figure 7.1: Notation describing the position of the observer, P, and of mass element dm.

Now, let us generalise these expression for the gravitational field due to an arbitrary body (Fig. 7.1). Denote by P the position of the observer, and by dm the mass element of the body. Their radius-vectors will be denoted by $\mathbf{r}$ and $\mathbf{r}'$, respectively:

$$\begin{aligned} \mathbf{r} &= (x, y, z) = (r\sin\theta\cos\lambda, r\sin\theta\sin\lambda, r\cos\theta) , \\ \mathbf{r}' &= (x', y', z') = (r'\sin\theta'\cos\lambda', r'\sin\theta'\sin\lambda', r'\cos\theta') , \end{aligned} \tag{7.9}$$

where x, y, z and x', y', z' are the Cartesian coordinates, and $r, \theta, \lambda, r', \theta', \lambda'$ the spherical coordinates of these points. Using the *superposition principle*, the gravitational potential of the body as a whole may be expressed as

$$U(P) = G\iiint \frac{\mathrm{d}m}{R} , \tag{7.10}$$

where the integration is carried out over all the mass elements of the body. This integral belongs to the category of Stieltjes integrals. Hence, the most general expression for the gravitational potential is given by Stieltjes integral (7.10).

However, we shall usually assume that the mass distribution within the body is sufficiently continuous to be able to introduce density, ρ. Then

$$\mathrm{d}m = \rho \mathrm{d}V \ , \tag{7.11}$$

where $\rho = \rho(x', y', z') = \rho(r', \theta', \lambda')$ and dV is the corresponding volume element. In this case, integral (7.10) simplifies to usual integral

$$U(P) = G\iiint_V \frac{\rho}{R}\mathrm{d}V \ , \tag{7.12}$$

where the integration is carried out over the volume of the body, V, i.e. over primed coordinates x', y', z', or r', θ', λ'. We shall usually express the gravitational potential in the general form (7.12). (Note that the density cannot be introduced at some discontinuities within the Earth, since the limits on the opposite sides are different.)

7.2 Gravity Field

Let us now describe the gravitational field of an arbitrary body in a rotation system.

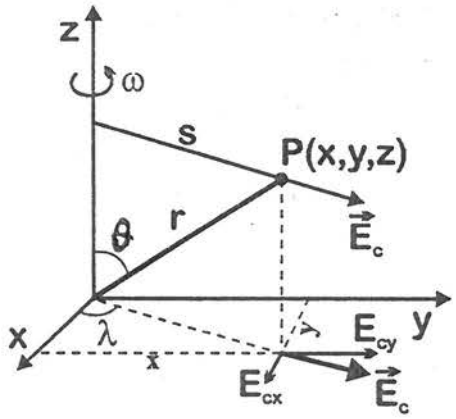

Figure 7.2: Decomposition of the centrifugal acceleration.

Assume that the body and the observer are fixed in a reference system which rotates about the z-axis at angular velocity ω (Fig. 7.2). The centrifugal acceleration at point P is $E_c = \omega^2 s$, where s is the distance from the rotation axis. From the similarity of the triangles in the xy-plane, it follows that

$$\frac{E_{cx}}{E_c} = \frac{x}{s} \ , \quad \frac{E_{cy}}{E_c} = \frac{y}{s} \ ,$$

where E_{cx} and E_{cy} are the components of E_c along axes x and y, respectively. By substituting $E_c = \omega^2 s$, we arrive at

$$\mathbf{E}_c = \left(\omega^2 x, \omega^2 y, 0\right) . \tag{7.13}$$

We are seeking the potential U_c of the centrifugal force, satisfying the relation

$$\mathbf{E}_c = \operatorname{grad} U_c = \left(\frac{\partial U_c}{\partial x}, \frac{\partial U_c}{\partial y}, \frac{\partial U_c}{\partial z}\right) . \tag{7.14}$$

By comparing (7.13) and (7.14), we arrive at the potential of the centrifugal force,

$$U_c = \frac{1}{2}\omega^2\left(x^2 + y^2\right) = \frac{1}{2}\omega^2 r^2 \sin^2\theta . \tag{7.15}$$

The force acting at a fixed point P of a rotating system is the superposition of the gravitational and centrifugal forces. We have thus arrived at the concept of the *gravity field* as the superposition of the gravitational field and the field of the centrifugal force. Hence, we shall refer to the vector sum of the gravitational and centrifugal accelerations as the *gravity acceleration*. Analogously, the *gravity potential* is the sum of the gravitational and centrifugal potentials,

$$U(P) = G\iiint_V \frac{\rho}{R}\mathrm{d}V + \frac{1}{2}\omega^2\left(x^2 + y^2\right) . \tag{7.16}$$

This is the basic, general formula for the gravity potential.

7.3 Expansion of the External Gravity Potential into a Series of Spherical Harmonics

By substituting the expansion (6.24a) for $1/R$ into (7.16) and using spherical coordinates, we get

$$U(r,\theta,\lambda) = \frac{G}{r}\iiint_V \rho \sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n P_n(\cos\gamma)\mathrm{d}V + \frac{1}{2}\omega^2 r^2 \sin^2\theta , \tag{7.17}$$

where we integrate over the points of volume V, i.e. over coordinates r', θ' and λ'.

Assume that the distance r of the observer from origin O is so large that $r > r'$ for all points of the body (the observer is outside the sphere which has its centre at origin O and contains the whole body in its interior). The series in

(7.17) is then convergent independently of coordinates r', θ', λ' (uniform convergence) and the integration and summation may be interchanged. Then

$$U(r, \theta, \lambda) = \sum_{n=0}^{\infty} \frac{Y_n(\theta, \lambda)}{r^{n+1}} + \frac{1}{2}\omega^2 r^2 \sin^2 \theta \ , \tag{7.18}$$

where

$$Y_n(\theta, \lambda) = G\iiint (r')^n P_n(\cos\gamma)\, \mathrm{d}m \ , \tag{7.19}$$

and dm being used instead of $\rho \mathrm{d}V$.

Let us investigate a few low terms of the series. The first spherical function Y_0 has a very simple physical meaning:

$$Y_0 = G\iiint \mathrm{d}m = GM \ , \tag{7.20}$$

where M is the total mass of the body.

To simplify the following terms, let us choose the coordinate system as follows:

1. The origin of the coordinate system, O, is at the centre of mass of the body, T. Consequently, the radius-vector of this centre is zero, i.e.

$$\mathbf{r}_T = \frac{1}{M}\iiint \mathbf{r}'\, \mathrm{d}m = 0 \ . \tag{7.21}$$

2. The axes of the coordinate system coincide with the principal axes of inertia. In this case, the deviatoric moments vanish.

Under these assumptions we shall calculate the other terms in series (7.18). For $n = 1$,

$$Y_1 = G\iiint r' \cos\gamma \, \mathrm{d}m = G\iiint r'[\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\lambda - \lambda')]\mathrm{d}m \ ,$$

where we have substituted (5.37) for cosγ. Thus

$$Y_1 = G\Big\{\cos\theta\iiint r'\cos\theta'\, \mathrm{d}m + \sin\theta\cos\lambda\iiint r'\sin\theta'\cos\lambda'\, \mathrm{d}m +$$

$$+ \sin\theta\sin\lambda\iiint r'\sin\theta'\sin\lambda'\, \mathrm{d}m\Big\} \ .$$

The first integral in this formula, after dividing by M, represents the z-coordinate of the centre of mass. However, according to Assumption 1, this coordinate is equal to zero. Similarly, the remaining two integrals correspond to the remaining

two coordinates of the centre of mass, and are also equal to zero. Consequently, under Assumption 1,

$$Y_1(\theta, \lambda) = 0 \ . \tag{7.22}$$

Since Y_1/r^2 represents the potential of a dipole, this means that the dipole component of the gravitational potential vanishes if the origin of the coordinate system is put at the centre of mass.

We shall see that the coefficients in function Y_2 can be expressed in terms of the moments of inertia of the body. We have had (see Section 6.9)

$$P_2(\cos\gamma) = \left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)\left(\frac{3}{2}\cos^2\theta' - \frac{1}{2}\right) + \tag{a}$$

$$+ 3\sin\theta\cos\theta\sin\theta'\cos\theta'\cos\lambda\cos\lambda' + \tag{b}$$

$$+ 3\sin\theta\cos\theta\sin\theta'\cos\theta'\sin\lambda\sin\lambda' + \tag{c}$$

$$+ \frac{3}{4}\sin^2\theta\sin^2\theta'\cos 2\lambda\cos 2\lambda' + \tag{d}$$

$$+ \frac{3}{4}\sin^2\theta\sin^2\theta'\sin 2\lambda\sin 2\lambda' \ , \tag{e}$$

where the individual parts of this formula have been denoted by the letters a to e.

It follows from Assumption 2 that the deviatoric moments vanish:

$$0 = J_{xy} = \iiint x'y' \, \mathrm{d}m = \iiint r'^2 \sin^2\theta' \cos\lambda' \sin\lambda' \, \mathrm{d}m \ ,$$

$$0 = J_{xz} = \iiint x'z' \, \mathrm{d}m = \iiint r'^2 \sin\theta' \cos\theta' \cos\lambda' \, \mathrm{d}m \ ,$$

$$0 = J_{yz} = \iiint y'z' \, \mathrm{d}m = \iiint r'^2 \sin\theta' \cos\theta' \sin\lambda' \, \mathrm{d}m \ .$$

By comparing these integrands with the expressions in e, b, c, it can be seen that the contributions of these parts to Y_2 will vanish.

It remains to express the contributions to Y_2 coming from parts a and d. Denote by A, B, C the moments of inertia about axes x, y, z, respectively. Then

$$A = \iiint \left(y'^2 + z'^2\right) \mathrm{d}m = \iiint r'^2 \left(\sin^2\theta' \sin^2\lambda' + \cos^2\theta'\right) \mathrm{d}m \ ,$$

$$B = \iiint \left(x'^2 + z'^2\right) \mathrm{d}m = \iiint r'^2 \left(\sin^2\theta' \cos^2\lambda' + \cos^2\theta'\right) \mathrm{d}m \ ,$$

$$C = \iiint \left(x'^2 + y'^2\right) \mathrm{d}m = \iiint r'^2 \sin^2\theta' \, \mathrm{d}m = \iiint r'^2 \left(1 - \cos^2\theta'\right) \mathrm{d}m \ .$$

Consequently,

$$A + B = \iiint r'^2 \left(\sin^2\theta' + 2\cos^2\theta'\right) \mathrm{d}m = \iiint r'^2 \left(1 + \cos^2\theta'\right) \mathrm{d}m \ ,$$

$$\frac{A+B}{2} - C = \iiint r'^2 \left(\frac{3}{2}\cos^2\theta' - \frac{1}{2}\right) \mathrm{d}m \ ;$$

see the expression in part *a*. The difference

$$B - A = \iiint r'^2 \sin^2\theta' \left(\cos^2\lambda' - \sin^2\lambda'\right) \mathrm{d}m$$

yields the contribution coming from part *d*. Hence,

$$Y_2 = G\left(\frac{A+B}{2} - C\right)\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) + G\frac{3}{4}(B - A)\sin^2\theta\cos 2\lambda \ . \qquad (7.23)$$

The gravity potential (7.18) may now be expressed as

$$U(r, \theta, \lambda) = \frac{GM}{r} + \frac{G}{2r^3}\left(C - \frac{A+B}{2}\right)\left(1 - 3\cos^2\theta\right) + \\ + \frac{3}{4}\frac{G}{r^3}(B - A)\sin^2\theta\cos 2\lambda + \frac{1}{2}\omega^2 r^2 \sin^2\theta + T(r, \theta, \lambda) \ , \qquad (7.24)$$

where

$$T(r, \theta, \lambda) = \sum_{n=3}^{\infty} \frac{Y_n(\theta, \lambda)}{r^{n+1}} \ . \qquad (7.25)$$

Term T will be referred to as the *disturbing potential*.

Let us introduce geocentric latitude $\phi = 90° - \theta$, and measure angle λ from the Greenwich meridian, not from the principal axis. Denoting by λ_0 the longitude of the principal axis (i.e. $\lambda_0 = 14.9°\,\mathrm{W}$), we must replace λ in (7.24) by $(\lambda - \lambda_0)$:

$$U(r,\phi,\lambda)=\frac{GM}{r}+\frac{G}{2r^3}\left(C-\frac{A+B}{2}\right)\left(1-3\sin^2\phi\right)+$$

$$+\frac{3}{4}\frac{G}{r^3}(B-A)\cos^2\phi\cos2(\lambda-\lambda_0)+\frac{1}{2}\omega^2r^2\cos^2\phi+T(r,\phi,\lambda)\ . \tag{7.26}$$

The first term on the right-hand side of the latter formula, i.e. GM/r, yields the potential of a spherically symmetric body. This is the main term in the expansion of the Earth's gravity field. We shall see that the remaining terms are three or more orders smaller. The second term contains the difference between the polar and equatorial moments of inertia, so that this term may be expected to be closely related to the flattening of the Earth; we shall prove this in Section 7.4. The next term, containing the difference $B-A$, describes the influence of the equatorial flattening. However, in the case of the Earth, this term is much smaller than the preceding one and, therefore, this term is very often neglected. The fourth term describes the centrifugal potential, and term T describes the influence of other irregularities in the density distribution of the body.

To simplify formula (7.26), let us assume that $B=A$ (actually, we may consider different values of A and B, but we transfer the corresponding small term to disturbing potential T). Then

$$\boxed{U(r,\phi,\lambda)=\frac{GM}{r}+\frac{G}{2r^3}(C-A)\left(1-3\sin^2\phi\right)+\frac{1}{2}\omega^2r^2\cos^2\phi+T(r,\phi,\lambda)}\ . \tag{7.27}$$

This form of the gravity potential is used in many applications. In addition to the leading term, GM/r, it explicitly contains the contributions due to the polar flattening and the non-inertiality of the reference system, whereas the other irregularities are formally hidden in disturbing term T. Formally, we shall also express formula (7.27) as

$$U(r,\phi,\lambda)=V(r,\phi)+T(r,\phi,\lambda)\ , \tag{7.28}$$

where term V represents the first three terms on the right-hand side of (7.27). We shall refer to term V as the *normal potential* or regular potential.

In some applications, e.g., in satellite geodesy, we need explicit expressions also for disturbing potential T. This is feasible if the addition theorem for Legendre polynomials is used in all terms Y_n, see (7.19) and (6.30):

$$Y_n(\theta,\lambda)=\sum_{m=0}^{n}\left[A_n^m\cos m\lambda+B_n^m\sin m\lambda\right]P_n^m(\cos\theta)\ , \tag{7.29}$$

where

$$A_n^m = G \in_m \frac{(n-m)!}{(n+m)!} \iiint (r')^n P_n^m(\cos\theta') \cos m\lambda' \, \mathrm{d}m \ ,$$

$$B_n^m = G \in_m \frac{(n-m)!}{(n+m)!} \iiint (r')^n P_n^m(\cos\theta') \sin m\lambda' \, \mathrm{d}m \ , \tag{7.30}$$

and $\in_0 = 1, \in_1 = \in_2 = \ldots = 2$. If the density distribution within the body were known, we could determine coefficients A_n^m and B_n^m by calculating integrals (7.30). This is not the case of the real Earth. However, these coefficients can be determined directly from perturbations of satellite orbits and, partly, from gravity measurements on the surface of the Earth.

Considering the leading role of the first term, $A_0^0 = GM$, and the choice of the origin of the coordinate system at the mass centre, gravity potential (7.18) may also be expressed as

$$U(r,\theta,\lambda) = \frac{GM}{r}\left\{1 + \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n}\left[J_n^m \cos m\lambda + S_n^m \sin m\lambda\right] P_n^m(\cos\theta)\right\} +$$

$$+\frac{1}{2}\omega^2 r^2 \sin^2\theta \ , \tag{7.31}$$

where a is a length factor, usually the equatorial radius of the Earth. In this case, coefficients J_n^m and S_n^m are dimensionless,

$$J_n^m = \frac{A_n^m}{GMa^n} \ , \quad S_n^m = \frac{B_n^m}{GMa^n} \ .$$

Note that several other variants of (7.31) are also used in the geophysical literature.

7.4 Equipotential Surfaces. The Geoid and Spheroid

7.4.1 Equation of the geoid

The expressions in the previous section make it possible to define the figure of the Earth physically as a figure of a selected equipotential surface of the gravity field. Note that, in this situation, Gauss and Bessel spoke of the mathematical figure of the Earth; see Chapter 1.

The equation

$$U(r,\phi,\lambda) = c_0$$

determines a system of equipotential surfaces of the gravity field if constant c_0 takes different values. The most important among these surfaces is the equipotential surface which is closest to the mean level of the seas. This equipotential surface is called the *geoid*. By mean sea level we understand the level from which all fluctuations have been removed, i.e. the fluctuations due to waves, tides and some other sources. We shall denote the corresponding constant for the geoid by U_0. Thus, the equation

$$\boxed{U(r, \phi, \lambda) = U_0} \tag{7.32}$$

represents the *equation of the geoid*. A recent value of the geoidal potential is (Burša et al., 1995)

$$U_0 = (62\,636\,857.0 \pm 1.0)\,\mathrm{m}^2\,\mathrm{s}^{-2}\ . \tag{7.33}$$

From the physical point of view, we shall take the figure of the geoid to be the figure of the Earth. In other words, by the figure of the Earth we shall understand the figure of the geoid.

Note that the gradient of the potential yields the gravity acceleration, **g**:

$$\mathbf{g} = \mathrm{grad}\, U\ . \tag{7.34}$$

7.4.2 Equation of the spheroid

The figure of the geoid is very complicated and, therefore, cannot be simply expressed mathematically. Hence, we shall approximate the geoid by the *spheroid*, by neglecting the disturbing term. Thus, using the notation in (7.28), the *equation of the spheroid* can formally be expressed as

$$\boxed{V(r, \phi, \lambda) = U_0}\,, \tag{7.35}$$

where U_0 is the same constant as in (7.32). The corresponding acceleration

$$\gamma = \mathrm{grad}\, V \tag{7.36}$$

is called the *normal gravity acceleration* or *normal gravity*. Note that the spheroid is not an equipotential surface of the gravity field, but only of its normal (regular) part.

Other approximations of the geoid are also used, such as the tri-axial ellipsoid, ellipsoid of revolution, or a sphere as the simplest approximation.

We shall now derive an explicit form of the equation of the spheroid (7.35). Using (7.27) the equation of the spheroid may be expressed as

$$U_0 = V = \frac{GM}{r}\left[1+\frac{1}{2r^2}\frac{C-A}{M}\left(1-3\sin^2\phi\right)+\frac{1}{2}\frac{\omega^2 r^3}{GM}\cos^2\phi\right]. \qquad (7.37)$$

After multiplying by r/U_0,

$$r = \frac{GM}{U_0}\left[1+\frac{K}{2r^2}\left(1-3\sin^2\phi\right)+\frac{\omega^2 r^3}{2GM}\left(1-\sin^2\phi\right)\right], \qquad (7.38)$$

where we have introduced $K = (C-A)/M$. This equation is the exact equation of the spheroid. For example, for a given value of ϕ we could use it to compute coordinate r. However, by multiplying this equation by r^2, we arrive at an algebraic equation of the 5th degree. Such an equation cannot be solved analytically, but only numerically. We could proceed as follows.

Let us assume for a moment that the numerical values of all constants in equation (7.38) are known. For a given value of ϕ, this equation then takes the form which is convenient for solving by the iterative method. We start the solution with a suitable choice of an initial value of the radial distance, r_0. For example, we can put $r_0 = a$ where a is the equatorial radius. By substituting this value into the right-hand side of equation (7.38), on the left-hand side we obtain a new value, r_1. By substituting r_1 again into the right-hand side, we obtain r_2, etc. Generally, this process can be described by the recurrence formula

$$r_{k+1} = f(r_k), \quad k = 0, 1, 2, \ldots,$$

where function f denotes the right-hand side of (7.38). It can be proved that this process will converge if $|\partial f/\partial r| < 1$ (if the derivative of the right-hand side, in its absolute value, is smaller than the derivative of the left-hand side). If r is close to the Earth's radius, e.g., to equatorial radius a, this derivative is very small (it is of the order of the Earth's flattening). Consequently, the corresponding iterative process will converge rapidly, and even the first iteration will be quite sufficient for many purposes.

Therefore, let us solve equation (7.38) approximately, retaining only the terms of the order of flattening α, and neglecting terms of the order of α^2 and of higher orders. By putting $r = a$ in the small terms on the right-hand side of (7.38) we arrive at

$$r = \frac{GM}{U_0}\left[1+\frac{J_2}{2}+\frac{q}{2}-\left(\frac{3}{2}J_2+\frac{q}{2}\right)\sin^2\phi\right], \qquad (7.39)$$

where we have introduced two dimensionless constants,

$$J_2 = \frac{K}{a^2} = \frac{C-A}{Ma^2}\ , \quad q = \frac{\omega^2 a^3}{GM}\ . \tag{7.40}$$

Parameter J_2 (Stoke's coefficient) represents one of the fundamental constant of the Earth. Its numerical value has been determined with a high accuracy from the perturbations of satellite motions. Parameter q has a simple physical meaning, because it can be expressed as

$$q = \frac{\omega^2 a}{GM/a^2} = \frac{\text{centrifugal acceleration at the equator}}{\text{gravitational acceleration}}\ .$$

The approximate values of these constants are as follows: $J_2 = 0.001$; $q = 0.003$. For further details we refer the reader to the end of Chapter 1.

Within the required accuracy, equation (7.39) may be modified to read

$$r = \frac{GM}{U_0}\left(1 + \frac{J_2}{2} + \frac{q}{2}\right)\left[1 - \left(\frac{3}{2}J_2 + \frac{q}{2}\right)\sin^2\phi\right]\ . \tag{7.41}$$

The latter equation is not exactly equivalent to (7.39), but it differs by a product of two small terms, which we neglect. Therefore, with an accuracy of the order of flattening, we consider equations (7.38), (7.39) and (7.41) as equivalent. Denote

$$a = \frac{GM}{U_0}\left(1 + \frac{J_2}{2} + \frac{q}{2}\right)\ , \tag{7.42}$$

$$\boxed{\alpha = \frac{3}{2}J_2 + \frac{q}{2}}\ . \tag{7.43}$$

Equation (7.41) can then be expressed as

$$\boxed{r = a\left(1 - \alpha\sin^2\phi\right)}\ . \tag{7.44}$$

The latter equation is also called *the equation of the spheroid*. It differs from the exact equation (7.38) by the squares of small terms of the order of flattening.

It is evident that parameter a, given by (7.42), is the equatorial radius. To determine the meaning of parameter α, let us denote the polar radius by c and substitute $\phi = 90°$ into (7.44):

$$c = a(1-\alpha) = a - a\alpha\ .$$

Hence,

$$\alpha = \frac{a-c}{a} ,$$

which means that parameter α is *the flattening of the Earth*.

Formulae (7.42) and (7.43) relate the geometric parameters of the Earth, i.e. its equatorial radius and flattening, to the physical parameters which characterise the gravity field of the Earth. For example, formula (7.43) can be used to compute the flattening if J_2 is determined from satellite observations and q by means of formula (7.40). Another physical method of determining the Earth's flattening will be described further on in this chapter.

It follows from the equation of the spheroid (7.44) that the spheroid is an axially symmetric surface, flattened at the poles. Therefore, it appears that it could be close to an ellipsoid. We shall indeed prove that both surfaces are identical down to the terms of the order of the flattening. However, they differ in terms of higher orders.

The equation of an ellipse,

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

can be expressed in the polar coordinates as

$$\frac{r^2 \cos^2 \phi}{a^2} + \frac{r^2 \sin^2 \phi}{c^2} = 1 .$$

Since $c = a(1-\alpha)$,

$$r = a\left[1 - \sin^2 \phi + (1-\alpha)^{-2} \sin^2 \phi\right]^{-1/2} = a\left(1 - \alpha \sin^2 \phi + \ldots\right) . \qquad (7.45)$$

Note that the spheroid is convenient from the physical point of view (it was derived from the analysis of the gravity potential), whereas the ellipsoid is more convenient from the geodetic point of view.

7.5 Normal Gravity

We have defined the normal gravity acceleration (normal gravity) by (7.36), i.e.

$$\gamma = \operatorname{grad} V .$$

We shall decompose this vector into the radial and meridional components. However, if we take radial components to be positive downwards, we must put

$$\gamma_r = -\frac{\partial V}{\partial r}\ ,\quad \gamma_N = \frac{\partial V}{r\partial \phi}\ , \tag{7.46}$$

γ_r being the radial component of the normal gravity and γ_N the northern component. The magnitude of the total vector, $\gamma = |\boldsymbol{\gamma}|$,

$$\gamma = \sqrt{\gamma_r^2 + \gamma_N^2} = \gamma_r\sqrt{1+\left(\frac{\gamma_N}{\gamma_r}\right)^2} \doteq \gamma_r\left[1+\frac{1}{2}\left(\frac{\gamma_N}{\gamma_r}\right)^2 + \ldots\right]. \tag{7.47}$$

It is evident that ratio γ_N/γ_r is of the order of flattening α.

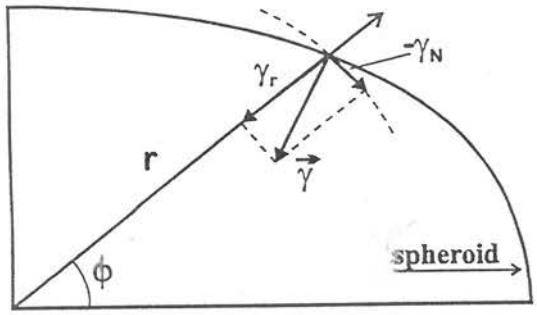

Figure 7.3: Decomposition of the normal gravity acceleration.

Let us again restrict ourselves to the terms of the order of the flattening and neglect the higher-order terms. Then

$$\gamma = \gamma_r\ , \tag{7.48}$$

and by using formulae (7.27) and (7.46), we arrive at

$$\boxed{\gamma = \frac{GM}{r^2} + \frac{3G(C-A)}{2r^4}\left(1-3\sin^2\phi\right) - \omega^2 r\left(1-\sin^2\phi\right)}\ . \tag{7.49}$$

This formula is the general formula for the external normal gravity. It contains not only the principal term, GM/r^2, but also the terms due to the flattening and the centrifugal acceleration. The formula may be used in many applications where the principal term on its own is not sufficient.

We are especially interested in the normal gravity on the spheroid. To obtain this function, we must substitute the equation of the spheroid (7.44) into the general formula (7.49). From the equation of the spheroid we get

$$\frac{1}{r^n} = \frac{1+n\alpha\sin^2\phi}{a^n}\ ,$$

the higher-order terms being omitted. Therefore, we shall substitute

$$\frac{1}{r^2} = \frac{1+2\alpha\sin^2\phi}{a^2}$$

in the principal term GM/r^2, but only $r = a$ in the remaining small terms:

$$\gamma = \frac{GM}{a^2}\left(1+2\alpha\sin^2\phi\right)+\frac{3G(C-A)}{2a^4}\left(1-3\sin^2\phi\right)-\omega^2 a\left(1-\sin^2\phi\right) =$$

$$= \frac{GM}{a^2}\left[1+\frac{3}{2}J_2 - q + \left(2\alpha - \frac{9}{2}J_2 + q\right)\sin^2\phi\right] \doteq$$

$$\doteq \frac{GM}{a^2}\left(1+\frac{3}{2}J_2 - q\right)\left[1+\left(2\alpha - \frac{9}{2}J_2 + q\right)\sin^2\phi\right],$$

where we have added a second-order term in the last line. We arrive at the final formula

$$\boxed{\gamma = \gamma_e\left(1+\beta\sin^2\phi\right)}, \tag{7.50}$$

where

$$\gamma_e = \frac{GM}{a^2}\left(1+\frac{3}{2}J_2 - q\right),$$

$$\beta = 2\alpha - \frac{9}{2}J_2 + q, \tag{7.51}$$

γ_e denoting the equatorial value of γ. Note that J_2, q and α were defined in the previous section by formulae (7.40) and (7.43). Formula (7.50) is called the *formula for the normal gravity*. It describes the normal gravity acceleration on the spheroid. (Quantity ϕ, which was introduced as the geocentric latitude, may be taken as the geodetic latitude in the first-order theory.)

Since the spheroid represents an approximation of the geoid, the formula for the normal gravity also describes a smoothed distribution of the gravity acceleration on the geoid. This means that we can determine constants γ_e and β in the normal gravity formula (7.50) from gravity measurements carried out on the surface of the Earth. Assume that such gravity measurements have been carried out at a large number of points on the surface of the Earth, including points at different latitudes. We assume that the measurements on the continents have been reduced to sea level; methods of reducing the gravity measurements will be described later. Denote by g_i the measured (or reduced) value of the gravity acceleration at the i-th point of latitude ϕ_i, and by γ_i the corresponding

normal gravity. Constants γ_e and β can then be computed, for example, by the method of least squares, requiring the sum of the squares

$$\sum_{i=1}^{N}(g_i-\gamma_i)^2$$

to be minimum.

The normal gravity formula cannot be used for geodetic purposes in the form (7.50), because terms of the order of α^2 have been neglected. Therefore, another term is considered in the normal gravity formula, which follows from the second-order theory:

$$\gamma=\gamma_e\left(1+\beta\sin^2\phi-\beta'\sin^2 2\phi\right), \tag{7.52}$$

where ϕ is the geodetic latitude. However, constant β', being very small, cannot be deduced from the observed values of gravity. It is determined theoretically. For example, if the spheroid is replaced by an ellipsoid of revolution, then $\beta'=5.9\times10^{-6}$; see Garland (1965).

7.6 Several Formulae for Normal Gravity

In practice, the constants in the normal gravity formula are not determined according to the latest measurements, but are given by agreement between countries for a given gravimetric system. In the countries of eastern Europe, Helmert's formula from the years 1909-1915 is still in use:

$$\gamma=978\,030\left(1+0.005\,302\sin^2\phi-0.000\,007\sin^2 2\phi\right), \tag{7.53}$$

where γ is given in milligals $\left(1\ \text{mgal}=10^{-3}\ \text{cm s}^{-2}=10\mu\text{m s}^{-2}\right)$. In many countries of western Europe, Cassinis' formula, which was adopted as the International Formula by the International Union of Geodesy and Geophysics in 1930, is used:

$$\gamma=978\,049\left(1+0.005\,288\,4\sin^2\phi-0.000\,005\,9\sin^2 2\phi\right). \tag{7.54}$$

Investigations of satellite orbits and some astronomical and geodetic measurements revealed that the parameters of the International Ellipsoid, adopted in 1924, and the parameters in the International Formula (7.54) differ considerably from reality. Therefore, a new reference ellipsoid and gravity formula were adopted in 1967. The gravity formula had the form

$$\gamma = 978\,031.8\left(1 + 0.005\,302\,4\sin^2\phi - 0.000\,005\,8\sin^2 2\phi\right) . \qquad (7.55)$$

For the first time, the value of γ_e in this formula was not determined from surface gravity data but from improved values of a, GM and J_2; see formula (7.51). Satellite data also formed the basis for the present Geodetic Reference System 1980 (IUGG, Canberra 1979). The gravity formula in this system has the form (Pick, 1987)

$$\gamma = 978\,032.677\,15\left(1 + 0.005\,279\,041\,4\sin^2\phi + \right.$$
$$+ 0.000\,023\,271\,8\sin^4\phi + 0.000\,000\,126\,2\sin^6\phi + \qquad (7.56)$$
$$\left. + 0.000\,000\,000\,7\sin^8\phi\right) .$$

For computing the changes of normal gravity with height above sea level, the following gravity formula is used in some countries (Pick, 1987):

$$\gamma = \gamma_0 - 0.308\,55h + 0.000\,000\,072h^2 - 0.000\,219h\cos 2\phi , \qquad (7.57)$$

where height h is given in metres and γ in milligals, γ_0 is the normal gravity on the reference ellipsoid. The last two terms in the latter formula are usually omitted. The formulae for the changes of the normal gravity with height are needed in computing gravity reductions, and in determining heights above sea level (see below).

7.7 Clairaut's Theorem

Clairaut (1743) proved that the flattening of the spheroid and the uniform variations of sea-level gravity from equator to poles were related.

Let us calculate the sum of flattening α and of coefficient β in the normal gravity formula. Using (7.43) and (7.51),

$$\boxed{\alpha + \beta = \frac{5}{2}q} , \qquad (7.58)$$

where q is given by (7.40). The latter relation is called *Clairaut's theorem*. Since parameter q can be determined from its definitions (7.40), and parameter β from gravity measurements, Clairaut's theorem enables us to determine flattening α. This is a physical (gravimetrical) method of determining the flattening of the Earth. Since the gravity measurements can be carried out also on the seas, this value of the flattening characterises the whole Earth. It represents the main

advantage over arc measurements, which can only be carried out on the continents.

Clairaut's theorem in the form of (7.58) is valid only to the terms of the first order of the flattening. For geodetic purposes it is again necessary to include the second-order terms (Heiskanen and Vening Meisnesz, 1958; Garland, 1965).

7.8 Methods of Determining the Flattening of the Earth

We should recapitulate the possibilities of determining the Earth's flattening. In principle, there are three methods of determining the flattening:

1. geometric method based on *arc measurements*;
2. physical method based on *Clairaut's theorem*;
3. astronomical (satellite) method, using *perturbations of satellite orbits* (their deviations from Kepler's laws).

The satellite method will be described in a greater detail in Chapter 12. Let us only note that it is possible to determine immediately the second-degree zonal geopotential parameter J_2. The flattening can then be obtained from formula (7.43) or from similar formulae containing also higher-order terms.

Chapter 8

The Geoid

Assume that the figure of the spheroid is known, i.e. its equatorial radius and flattening α have already been determined. To determine the geoid it is now sufficient to determine the distances between the geoid and the spheroid. Therefore, in this chapter we shall deal with the deviations of the geoid from the spheroid. We shall frequently speak of *undulations of the geoid.*

In this chapter we shall restrict ourselves to the zero-order theory. Thus we shall neglect even terms of the order of the flattening, although such terms were considered in the previous chapter. This approximation is justified by the fact that the distances between the geoid and spheroid are relatively small, only exceptionally exceeding 100 m.

The determination of the geoid and related questions belonged to the most important problems of higher geodesy from the end of the 19th century up to the 1950's. The geoid is important mainly from a *practical* point of view. Namely, all surface measurements are carried out in the gravity field of the Earth and all measuring instruments are adjusted to their correct position by means of levels, i.e. with regard to the plane tangent to the local equipotential surface of the gravity field. The measured values are then reduced to "the sea level", i.e. to the level of the geoid. From the geoid they are then transferred to a surface which is defined mathematically, usually to an ellipsoid of revolution. Therefore, it is necessary to known the shape of the geoid and its position with respect to the mathematical surface (Pick, 1987). Nevertheless, certain problems in determining the geoid accurately led some authors to looking for other geodetic methods which did not require the geoid to be known. One of these methods is Molodenskii's method, which we shall describe later.

The introduction of satellites changed the situation dramatically. On the one hand, it diminished the importance of the geoid as a reference surface, because many problems of higher geodesy can now be solved in three-dimensional space without introducing auxiliary surfaces. On the other hand, satellite techniques considerably improved our knowledge of the geoid on the whole Earth. Of particular importance are the detailed data on the shape of the geoid on the oceans and seas, obtained by satellite altimetry techniques. They represent very promising material for future geophysical interpretations.

In this chapter we shall describe the main classical methods and relations allowing the undulations of the geoid to be determined. At the end of the chapter we shall also present some results of satellite investigations, but the corresponding theory will be described later.

8.1 Bruns' Theorem

It is evident from the definitions of the geoid and spheroid that the deviations between these surfaces are connected with the disturbing part of the gravity potential, T.

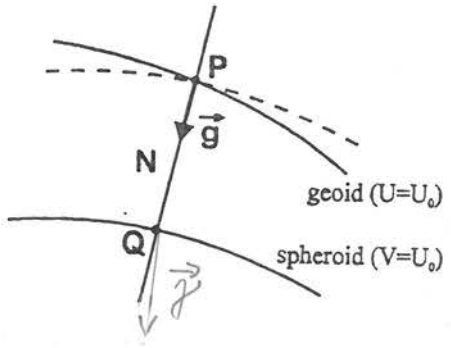

Figure 8.1: Geometry for deriving Bruns' theorem.

Consider a point P on the geoid, and the plumb line of the gravity field passing through this point (Fig. 8.1). Denote by Q the point of intersection of this plumb line with the spheroid. The undulation of the geoid, N, will be taken positive if the geoid is above the spheroid $(N > 0)$, and negative if the geoid is below the spheroid $(N < 0)$. The decomposition of the gravity potential U into the normal part V and the disturbing part T can be described as

$$U(P) = V(P) + T(P) \ . \tag{8.1}$$

The definitions of the geoid and spheroid yield

$$U(P) = U_0 = V(Q) \ . \tag{8.2}$$

By comparing the right-hand sides of (8.2) and (8.1) we obtain

$$V(Q) - V(P) = T(P) \ . \tag{8.3}$$

However,

$$V(Q) - V(P) \doteq \gamma(Q) N \cos \in \doteq \gamma(Q) N \ , \tag{8.4}$$

where $\in$ is the angle between the normals to the geoid and spheroid. Since we want to determine only the leading term contributing to distance N, we shall neglect the angle between the normals, i.e. we shall put $\in = 0$. As distance N does not usually exceed 100 m, we shall also neglect the change of γ between points P and Q, and put $\gamma = \gamma(Q)$. Consequently, we arrive at a simple formula

$$\boxed{N = \frac{T(P)}{\gamma(Q)}}, \tag{8.5}$$

which is called *Bruns' theorem*.

Distance N can now be determined by using Bruns' theorem, because disturbing potential T can be determined directly from satellite measurements (actually, the disturbing potential can only be approximated by a finite, but sufficiently long series, the coefficients of which can be determined by satellite methods). However, in the past it was not possible to determine the disturbing potential directly. Therefore, it was necessary to look for other methods, which we shall describe in the following sections.

8.2 Fundamental Equation of Physical Geodesy

Instead of potentials U and V from the preceding section, we shall now consider their gradients, i.e. accelerations $g(P)$ and $\gamma(Q)$.

Differentiating (8.1) with respect to the normal to the geoid, $n(P)$, yields approximately

$$g(P) = \gamma(P) + \frac{\partial T(P)}{\partial n(P)}, \tag{8.6}$$

where we have again assumed that $\in = 0$ in determining $\gamma(P)$. Since the normal gravity is usually given on the spheroid, we can put approximately

$$\gamma(P) = \gamma(Q) - \frac{\partial \gamma(Q)}{\partial n(Q)} N, \tag{8.7}$$

where $n(Q)$ is the normal to the spheroid at point Q. Using Bruns' theorem we arrive at

$$\boxed{\frac{\partial T(P)}{\partial n(P)} - \frac{T(P)}{\gamma(Q)} \frac{\partial \gamma(Q)}{\partial n(Q)} - \left[g(P) - \gamma(Q)\right] = 0}. \tag{8.8}$$

This equation is called the *fundamental equation of physical geodesy*. Quantity

$$\Delta g = g(P) - \gamma(Q) \tag{8.9}$$

is called *the apparent gravity anomaly*, as it relates the fields at two different points, P and Q. The true gravity anomaly at point P would be $g(P) - \gamma(P)$.

Equation (8.8) has the form of a differential equation for computing the disturbing potential T. However, the disturbing potential cannot be determined

solely from this equation, as the gravity anomalies are usually known on the geoid only, whereas the solution of differential equation (8.8) requires the anomalies to be known along the normal to this surface.

We usually assume that there are no masses outside the geoid (see the discussion of this problem in the chapter on gravity reductions). The disturbing potential outside the geoid is then a harmonic function and satisfies Laplace's equation

$$\Delta T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0 \ . \tag{8.10}$$

Equation (8.8) can then be considered a boundary condition in solving equation (8.10). As the boundary condition contains a linear combination of T and $\partial T/\partial n$, the determination of the disturbing potential by means of (8.10) and (8.8} represents the third boundary problem of the potential theory. However, a direct numerical solution of this problem, for example by means of the finite difference method, is complicated even with the help of computers. The situation in the past was even worse, so that an analytical solution to this problem was required. Such a solution was proposed by Stokes (1849); see Section 8.3.

8.3 Stokes' Theorem

To determine the disturbing potential $T(P)$, appearing in Bruns' theorem and in the fundamental equation of physical geodesy, we must determine the corresponding terms Y_n in the series of the gravity potential:

$$T(P) = \sum_{n=3}^{\infty} \frac{Y_n}{r^{n+1}} \ ; \tag{8.11}$$

see (7.25). Let us express the fundamental equation of physical geodesy in terms of Y_n and neglect the higher-order terms. We shall approximate the individual terms as follows:

$$\begin{gathered} \frac{\partial T}{\partial n} \doteq -\frac{\partial T}{\partial r} = \sum_{n=3}^{\infty} (n+1) \frac{Y_n}{r^{n+2}} \ , \\ \gamma \doteq \frac{GM}{r^2}, \quad \frac{\partial \gamma}{\partial n} \doteq -\frac{\partial \gamma}{\partial r} = \frac{2GM}{r^3}, \quad \frac{1}{\gamma}\frac{\partial \gamma}{\partial n} = \frac{2}{r} \ . \end{gathered} \tag{8.12}$$

The fundamental equation of physical geodesy (8.8) can now be expressed in the following simplified form

$$\sum_{n=3}^{\infty}(n-1)\frac{Y_n}{r^{n+2}} = \Delta g \ . \tag{8.13}$$

We wish to determine the individual terms Y_n from this equation. This can be done if we expand anomaly Δg on the right-hand side of this equation into a similar series.

Consider a harmonic function f, i.e. a function satisfying Laplace's equation, outside a sphere S of unit radius. Its value at a point with spherical coordinates θ, λ is related to its values at other points of the sphere by the formula

$$f(\theta,\lambda) = \sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\iint_S f(\theta',\lambda')P_n(\cos\psi)\,\mathrm{d}\sigma \ , \tag{8.14}$$

where $\mathrm{d}\sigma$ is a surface element, the integration is performed over the primed coordinates, and ψ is the angle between the radius-vectors of the points with coordinates θ, λ and θ', λ' (the angular distance of these points as seen from the centre of the sphere). For the proof of (8.14) we refer the reader to Heiskanen and Vening Meinesz (1958).

Approximating the geoid by a sphere (which introduces an error of the order of the flattening, which we shall neglect), and assuming that there is no mass outside the geoid, the gravity anomaly will be a harmonic function on the sphere, and formula (8.14) will apply to it. We obtain from (8.13)

$$\sum_{n=3}^{\infty}(n-1)\frac{Y_n}{r^{n+2}} = \sum_{n=0}^{\infty}\frac{2n+1}{4\pi}\iint_S \Delta g P_n(\cos\psi)\,\mathrm{d}\sigma \ , \tag{8.15}$$

where the integration is performed over the unit sphere. The left-hand side of the latter formula does not contain the spherical harmonics of degrees $n = 0, 1, 2$, the summation starts with $n = 3$. Consequently, the right-hand side cannot contain terms of low degrees either, and we may change the summation from $n = 3$. Namely, if the low-degree components of the gravity field are completely included in the normal field, the corresponding integrals on the right-hand side, i.e.

$$\iint_S \Delta g P_n(\cos\psi)\,\mathrm{d}\sigma$$

will be equal to zero for $n = 0, 1, 2$. We shall return to this problem later.

The sums on both sides of (8.15) will be equal to one another if the individual terms are equal. Multiplying each term by $r/(n-1)$ and summing yields

$$T=\sum_{n=3}^{\infty}\frac{Y_n}{r^{n+1}}=\frac{r}{4\pi}\sum_{n=3}^{\infty}\frac{2n+1}{n-1}\iint_S \Delta g P_n(\cos\psi)\,\mathrm{d}\sigma\ . \tag{8.16}$$

By exchanging the summation and integration, and using Bruns' theorem (8.5), we get

$$\boxed{N=\frac{r}{4\pi\gamma}\iint_S \Delta g S(\psi)\,\mathrm{d}\sigma}\ , \tag{8.17}$$

where

$$S(\psi)=\sum_{n=3}^{\infty}\frac{2n+1}{n-1}P_n(\cos\psi)\ . \tag{8.18}$$

In practice the term for $n=2$ is also included in the sum in (8.18). As we have mentioned, if the second-degree term of the gravity field is fully contained in the normal field, it will not change the integral in (8.17), because the corresponding contribution for $n=2$ will be equal to zero. However, we have omitted, for example, the term in the series of the gravity potential containing the difference of the equatorial moments of inertia, $B-A$. This term is of the second degree. Also the flattening of our selected ellipsoid may contain an error. For this reason it is reasonable to include the second-order term in (8.18). Consequently,

$$S(\psi)=\sum_{n=2}^{\infty}\frac{2n+1}{n-1}P_n(\cos\psi)\ . \tag{8.19}$$

The latter sum may be expressed in terms of elementary functions in the form (Heiskanen and Vening Meinesz, 1958)

$$S(\psi)=\frac{1}{\sin\frac{\psi}{2}}+1-5\cos\psi-6\sin\frac{\psi}{2}-3\cos\psi\ln\left(\sin\frac{\psi}{2}+\sin^2\frac{\psi}{2}\right)\ . \tag{8.20}$$

Formula (8.17) is called Stokes' formula, or *Stokes' theorem*. It enables us to calculate the distance N between the geoid and spheroid by means of gravity anomalies Δg. However, these anomalies should be known on the whole Earth, which represents a certain limitation of this important theorem. In this respect, especially the construction of accurate instruments for gravity measurements at sea extended the applicability of Stokes' theorem substantially (Vening Meinesz, 1929).

There are also other practical and theoretical problems in applying Stokes' theorem. Firstly, in order to determine the gravity anomaly Δg, defined by (8.9), it is necessary to know gravity acceleration g on the geoid. However, gravity measurements on the continents are carried out on the physical surface of the Earth above the geoid. Consequently, it is necessary to reduce the measured

values to expected geoidal values by introducing suitable gravity reductions (see Chapter 10). Secondly, in deriving Stokes' theorem we have assumed that there are no masses outside the geoid, i.e. above sea level. This requirement is not fulfilled on the continents either, and we must try to solve this problem again by means of suitable gravity reductions. Actually, we try to solve the problem for an auxiliary ("regularized") model of the Earth, which has the same geoid as the real Earth, but no masses above the geoid. We should also consider atmospheric masses, but it can be shown that their effect is very small and can be neglected (Heiskanen and Vening Meinesz, 1958).

Stokes' theorem enabled the first maps of the undulations of the geoid to be constructed. The present maps also use extensive satellite data (further details are in this chapter below).

8.4 Vening Meinesz Formulae for the Deflections of the Vertical

By deflection of the vertical we understand the angle between the plumb line of the gravity field and the plumb line of the normal field at a particular point. Let us consider these deflections only at points on the geoid. Since these deflections are connected with the undulations of the geoid, we may expect the corresponding theory to be closely related to Stokes' theorem.

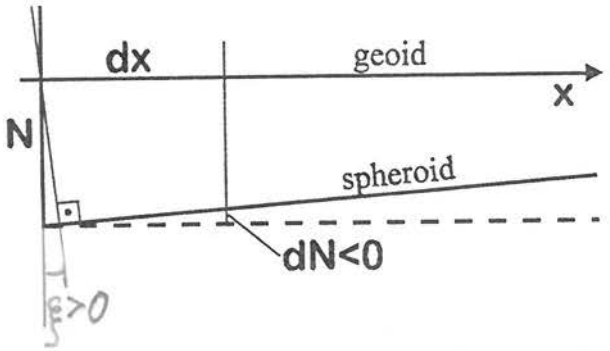

Figure 8.2: Relation between the undulations of the geoid and the deflections of the vertical.

We shall describe the total deflection in terms of its meridional (northern) and perpendicular (eastern) components, and denote them by ξ and η, respectively. These components can be expressed approximately as

$$\xi \doteq \tan\xi = -\frac{\partial N}{\partial x}, \quad \eta \doteq -\frac{\partial N}{\partial y}, \tag{8.21}$$

where the x-axis is oriented northward, and the y-axis eastward (Fig. 8.2).

Consider the neighbourhood of a point P on the geoid. A perturbing body at azimuth α inclines the vertical at P in the vertical plane determined by this azimuth. The corresponding contribution to the geoidal undulation has isolines which are shown as the dashed lines in Fig. 8.3. The figure shows that $\cos\alpha = (r\,\mathrm{d}\psi)/\mathrm{d}x$, i.e. $1/\mathrm{d}x = (\cos\alpha)/(r\,\mathrm{d}\psi)$ Similarly $1/\mathrm{d}y = (\sin\alpha)/(r\,\mathrm{d}\psi)$. By substituting these expressions into (8.21) we arrive at

$$\xi = -\frac{1}{4\pi\gamma}\iint_S (\cos\alpha)\frac{\mathrm{d}S(\psi)}{\mathrm{d}\psi}\Delta g\,\mathrm{d}\sigma\ ,$$

$$\eta = -\frac{1}{4\pi\gamma}\iint_S (\sin\alpha)\frac{\mathrm{d}S(\psi)}{\mathrm{d}\psi}\Delta g\,\mathrm{d}\sigma\ . \qquad (8.22)$$

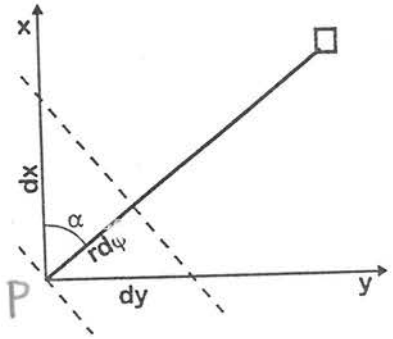

Figure 8.3: Geometry for deriving the Vening Meinesz formulae; details are described in the text.

Formulae (8.22) are called the *Vening Meinesz formulae* for the deflections of the vertical (Vening Meinesz, 1928). They play an important role in determining the position of points on the Earth's surface and in constructing world geodetic systems. The deflections of the vertical should be known at the initial point of each geodetic system. Maps of these deflections can also be used for geological interpretations, for example, in studies of the crustal structure or convective currents in the mantle. (However, a better possibility of studying mantle convection has been opened by investigations of seismic anisotropy).

8.5 Maps of the Geoid

Several examples of older maps of the undulations of the geoid, derived on the basis of Stokes' theorem, can be found in the monograph by Heiskanen and Vening Meinesz (1958). Of the later geoids we would like to mention the Columbus and Kaula geoids; for details and corresponding maps we refer the reader to Garland (1965). The Columbus geoid, computed at the Ohio State

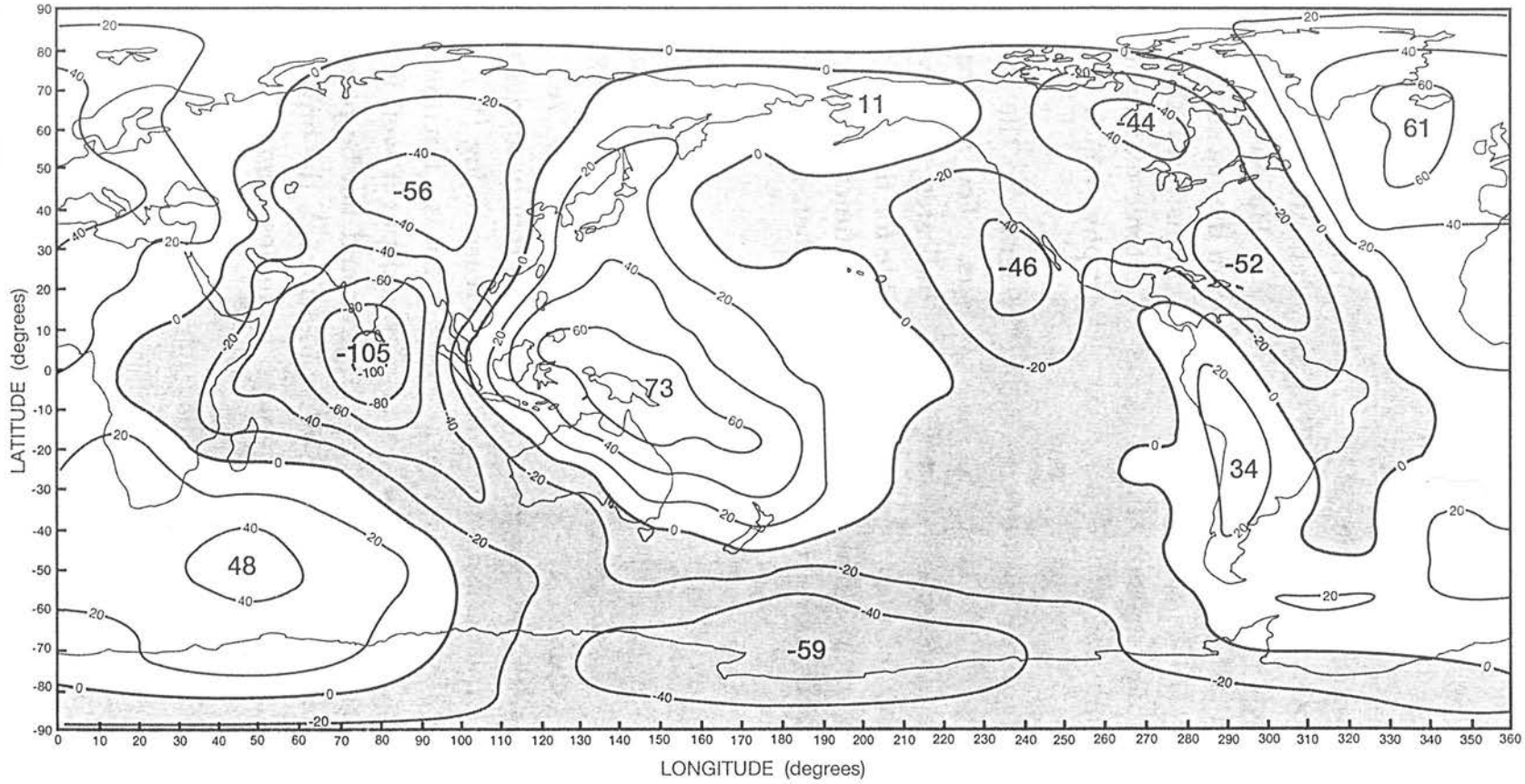

Figure 8.4: Contours of geoid height (in metres) relative to the reference ellipsoid. (After (Stacey, 1992)).

University in 1957, was based on isostatic anomalies (the term of "isostasy" will be explained in Chapter 9). In 1961 Kaula derived a geoid based on free-air gravity anomalies (their spherical harmonic expansion to order 8), astrogeodetic data and also on satellite data. These geoids were rather smooth, not detailed yet. The maximum undulations of the geoid on these maps were of about 50 m. However, these maps have already shown the important fact that the areas of positive or negative undulations of the geoid do not generally coincide with the distribution of the continents and oceans. This indicates that the continental masses must be compensated at greater depths.

Later determinations of the geoid made use of satellite data to a much greater extent. The most detailed information on the shape of the geoid on the oceans (the shape of the mean sea level) was obtained by means of the satellite altimetry. For details and maps we refer the reader to the reviews by Burša and Pec (1993), and Wahr (1996).

In Fig. 8.4 we reproduce the schematic map of the geoid from the textbook by Stacey (1992). The map shows clearly that the undulations of the geoid do not coincide with the distribution of the continents. It should also be mentioned that the largest deviation of the geoid from the ellipsoid occurs in the region of southern India (undulation of about -100 m, i.e. the geoid is below the ellipsoid). The geophysical interpretation of the maps of the geoid will also be discussed in Chapter 9.

Chapter 9

Isostasy

Maps of the geoid show that the undulations of the geoid do not correspond to the division of the Earth's surface into oceans and continents. If the masses of the continents were added on a laterally homogeneous ellipsoid, the isolines of the geoid would copy the contours of the continents. However, this is not so in the maps of the geoid. This means that the surplus of mass above the geoid must be compensated by a mass deficiency below the geoid. This phenomenon of compensation, which is analogous to hydrostatic equilibrium in liquids, is called *isostasy*. In other words, at a certain depth, the pressure is the same regardless to the position of the point below the continent, ocean, etc. This depth is called the *depth of compensation*.

The idea of compensation of mountains is not new. Speculations about it can be found in the papers of Leonardo da Vinci, and the first measurements indicating the existence of the isostasy were performed by Bouguer in about 1750 in Peru. Detailed investigations of isostasy started with the geodetic measurements in India around 1850. They are connected mainly with the interpretations of Pratt and Airy.

We shall mention a few details of these measurements. The famous geodesist Everest compared the geographical coordinates of two points, in Kaliana and Kalianpur, at different latitudes. The latitude difference between these points computed along a triangulation traverse (with respect to an ellipsoid) was found to be larger than that obtained from astronomical observations. *J. H. Pratt* explained this discrepancy by the deflection of the vertical due to the attraction of the Himalayas. Evidently the Himalayas deflected the vertical. However, when Pratt computed the deflection of the vertical caused by the topographic masses of the Himalayas, he found that the observed deflections represented only about 1/3 of the theoretical values. This result showed that 2/3 of the horizontal attraction of the Himalayas at these points had to be compensated by mass deficiencies in or under the mountain mass.

Two basic conceptions were proposed by J. H. Pratt and by G. B. Airy for explaining the mechanism of compensation. According to Pratt, the compensation is achieved by variations of density, i.e. the higher the mountain, the smaller its mean density. According to Airy, the mountains have similar mean densities, but higher mountains sink deeper into the underlying layer, having a higher density. Thus, Airy assumed the existence of "mountain roots". Some details of these two theories are given in the sections which follow.

9.1 Pratt-Hayford Isostatic System

This system is based on Pratt's ideas, but Hayford added to it a method for computing the corresponding isostatic reductions.

Denote by H the depth of compensation, and by ρ_0 the density of the column which reaches up to the sea level (Fig. 9.1). Another column, reaching to height h above sea level, is assumed to have a smaller density ρ. We wish to determine the density difference $\Delta\rho = \rho - \rho_0$ if H, ρ_0 and h are known.

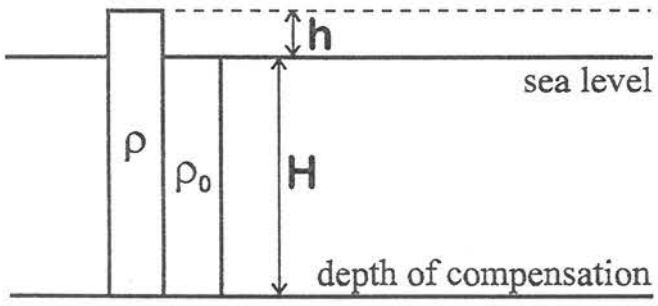

Figure 9.1: Pratt's isostatic hypothesis.

The condition of equality of pressure at the depth of compensation can be expressed approximately by the relation

$$\rho(H+h) = \rho_0 H \ . \tag{9.1}$$

Actually, the condition of hydrostatic equilibrium at depth H should be expressed as

$$\int_{-H}^{h} \rho g \, \mathrm{d}z = C \ ,$$

but we assume constant density ρ in each block, and ignore the vertical change of gravity acceleration g. Consequently, we shall restrict ourselves to the simplified relation (9.1).

Assuming $h << H$, relation (9.1) may be modified to read

$$\rho = \rho_0 \frac{H}{H+h} = \rho_0 \frac{1}{1+\dfrac{h}{H}} \doteq \left(1 - \frac{h}{H}\right)\rho_0 \ .$$

Hence, the density difference, $\Delta\rho = \rho - \rho_0$, comes out as

$$\boxed{\Delta\rho = -\frac{h}{H}\rho_0} \ . \tag{9.2}$$

It has become a tradition in gravimetry to take density $\rho_0 = 2{,}67\,\mathrm{g\,cm^{-3}}$. This value corresponds to the mean density of granitic rocks. Only one

parameter on the right-hand side of (9.2) is then unknown, namely compensation depth H. It is necessary to select different values of H and calculate the corresponding densities. For the density models obtained in this way, we can calculate the gravity field on the surface, and compare it with the observed field. For the depth of compensation we then adopt that value of H which yields the best agreement between the theoretical and observed gravity fields. For example, Hayford and Bowie adopted the value $H = 113.7$ km for the territory of the USA.

The columns in Fig. 9.1 have different heights. To simplify the computations of the gravity effect of these columns, Hayford measured the depth of compensation from the Earth's surface, not from the sea level. Consequently, the level of compensation under the continents is higher then under the oceans. This is a small deviation from the conception of Pratt. Formula (9.2) for computing $\Delta\rho$ was adopted by Hayford without changes. Now, when the computations are performed by computers, such technical simplifications are not necessary.

Pratt's model can be applied only rarely. It can be applied only in regions with a characteristic differentiation of masses, e.g., in thermally anomalous regions.

9.2 Airy-Heiskanen Isostatic System

Let us introduce a notation which differs a little from that used in the previous section (see Fig. 9.2): ρ_c is the density of the crust, the same in all columns; H is the thickness of the crust in the regions where the surface is at sea level; h is the height above sea level; h' is the submersion below the "normal" depth, i.e. the height of the mountain root; ρ_m is the density in the mantle.

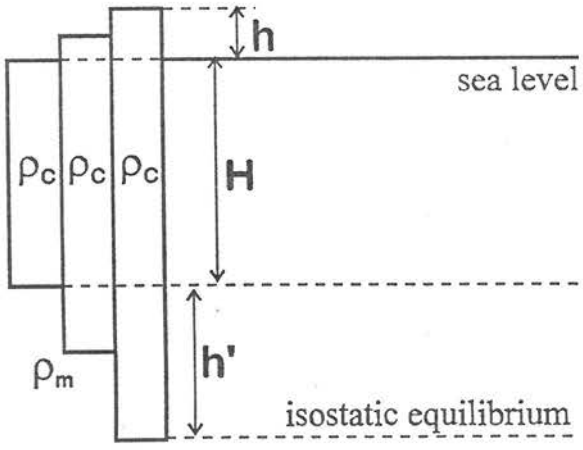

Figure 9.2: Airy's isostatic hypothesis.

The condition of isostatic equilibrium may be expressed as

$$\rho_c H + \rho_m h' = \rho_c (h + H + h') \ , \tag{9.3}$$

which yields

$$\boxed{h' = h\frac{\rho_c}{\rho_m - \rho_c}}\,. \tag{9.4}$$

The usual values are $\rho_c = 2..67\,\mathrm{g\,cm^{-3}}$ and $\rho_m = 3.27\,\mathrm{g\,cm^{-3}}$. Consequently,

$$h' \doteq 4.5h\,.$$

For example, if the Earth's crust at sea level is of thickness $H = 30$ km, then in mountains of height $h = 2$ km, the thickness of the crust should be $2+30+2\times4.5 = 41$ km. As opposed to the thick crust in mountain regions, the crust in the regions of oceans should be thin. This was confirmed by seismic measurements.

Airy's model of isostatic compensation usually describes the real situation better than Pratt's model. Nevertheless, both the models are too simple to be used in detailed investigations of the crustal and upper mantle structure. For such studies, seismic investigations are necessary.

9.3 Vening Meinesz Regional Isostatic System

The Pratt-Hayford and Airy-Heiskanen isostatic systems are local systems, i.e. they assume that the compensation occurs just below the topographic masses. However, the Earth's crust is solid, having a certain strength, so that the compensation cannot be quite local. Small topographic masses, such as individual hills or small mountain ranges, cannot be locally isostatically compensated according to the Pratt-Hayford or Airy-Heiskanen models. These facts were taken into account by Vening Meinesz, who modified Airy's hypothesis, and elaborated a system of regional isostatic compensation (Fig. 9.3). It includes the computation of the bending of an elastic plate.

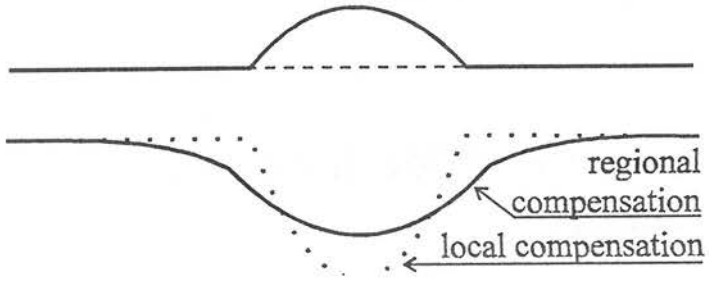

Figure 9.3: Vening Meinesz's regional compensation.

Chapter 10

Gravity Measurements and Their Reductions

10.1 Review of Methods of Gravity Measurements

In this section we shall briefly review the methods of gravity measurements. Very detailed descriptions of the classical gravity measurements can be found in Heiskanen and Vening Meinesz (1958), and in Pick et al. (1973). Also Garland (1965) devoted much attention to these problems. Instruments and techniques, presently used, were described in detail by Wahr (1996).

The methods of measuring the gravity acceleration for purposes of geodesy and geophysics can be subdivided schematically as follows (the methods which are now preferred are given in italics):

- absolute
  - pendulum
  - ballistic (*free fall*)
- relative
  - pendulum
  - *gravimeter* (spring system)
    - prospecting (~ 0.1 mgal)
    - geodetic (~ 0.01 mgal)

The gravity acceleration may, in principle, be measured by means of any process or phenomenon which is influenced by this acceleration. The best known of these processes, discussed in basic courses of physics, are the free fall of a body, and the motion of the mathematical pendulum. Both of these methods have been used in gravity measurements, but their technical realisation had to be modified with a view to the required accuracy. The accuracy now required of the best instruments for gravity measurements is very high, of about 0.01 mgal, i.e. $0.1\,\mu\mathrm{m\,s}^{-2}$. Considering that $g \doteq 10\,\mathrm{ms}^{-2}$, it represents a very high relative accuracy of about 10^{-8}.

The gravity acceleration also appears in the formula for the weight of a body, $W = mg$, where W is the weight, m the mass, and g the gravity acceleration. This principle of measuring the gravity acceleration by measuring the weight of a small body is used in gravimeters.

Another option is, for example, measuring the motion of a projectile, instead of which motions of satellites are studied (to reduce the problems with atmospheric drag). However, in this chapter we shall restrict ourselves to ground-based methods only. Satellite methods will be described later.

The *absolute gravity measurement*, i.e. the determination of the whole value of the gravity acceleration, represents a very difficult problem. Therefore, these measurements have been performed only at a very limited number of points on the Earth's surface. At other points, e.g., in gravimetric prospecting, only the

differences of the gravity acceleration are measured with respect to the point where the absolute value is known. In this case we speak of *relative gravity measurements*. We shall now give some details of and comment on these methods.

10.1.1 Absolute pendulum measurements

The well-known formula for the period of the mathematical pendulum is

$$T = 2\pi\sqrt{\frac{l}{g}} ,$$

where l is the length of the mathematical pendulum, and g is the gravity acceleration. However, this formula is valid only for infinitely small amplitudes of oscillations. For the purposes of absolute gravity measurements, a more accurate formula is required. The following simple formula is usually sufficient:

$$T = 2\pi\sqrt{\frac{l}{g}}\left(1+\frac{1}{16}\alpha^2\right) ,$$

α being the amplitude of oscillations; for the derivation of this formula we refer the reader to Heiskanen and Vening Meinesz (1958), and Pick et al. (1973). Term $\alpha^2/16$ represents a correction for an infinitely small amplitude. However, several further corrections must be introduced as well:

- correction in respect of vacuum;
- temperature corrections;
- correction for the geomagnetic field (Foucault currents);
- correction for the co-oscillation of the stand.

As a consequence of these technical problems with pendulum measurements, methods based on the free fall of a body are now preferred.

10.1.2 Free-fall methods

These methods require accurate measurements of distance and time. Present optical (laser interferometry) and electronic techniques fulfil these requirements. The free fall of a corner reflector in vacuum is usually measured in these methods. Here are several values obtained by these methods:

1) Potsdam (Germany), 1976:

$$g = \left(981\,261.424 \pm 0.010\right) \times 10^{-5}\,\mathrm{m\,s^{-2}} ,$$

which represents relative accuracy of 10^{-8}.

2) Paris, 1977:

$$g = (980\,926.019 \pm 0.015) \times 10^{-5}\,\mathrm{m\,s^{-2}}.$$

3) Pecný (Czech Republic), 1992:

$$g = (980\,933.274\,0 \pm 0.006\,6) \times 10^{-5}\,\mathrm{m\,s^{-2}}.$$

10.1.3 Relative measurements by means of gravimeters

The application of gravimeters simplifies gravity measurements in comparison with the techniques mentioned above, but some new problems appear, in particular:

- gravimeters, as all instruments for relative measurements, must be calibrated;
- the elastic properties of springs and fibres in the gravimeter vary with temperature and with time;
- some gravimeters are also sensitive to variations in atmospheric pressure and to vibrations;
- the scales of some gravimeters are non-linear.

To overcome these problems, special technical solutions and methods of measurement are used. For details we refer the reader to the books mentioned at the beginning of this chapter. Brief information about several gravimeters is given in Tab. 10.1.

Table 10.1: Basic characteristics of several gravimeters.

Gravimeter	Made in	Weight (kg)	Accuracy $(\mu\mathrm{m\,s^{-2}})$
Worden	USA	3	0.2 to 0.5
Sharpe	Canada	3	0.1
La Coste and Romberg	USA	5	0.1

10.2 Reductions of Gravity Measurements

To be able to apply the gravity measurements in geophysics and geodesy, it is necessary to reduce the measured values to one equipotential surface, usually to the geoid. We speak of reductions of gravity measurements. Many reductions have been proposed. A brief review of them follows:

a) Non-isostatic reductions:

- free-air reduction (Faye reduction, Faye's reduction),
- Bouguer reduction with terrain correction,
- Helmert condensation reductions,

- Rudzki inversion reduction.

b) Isostatic reductions:

- Pratt-Hayford reduction,
- Airy-Heiskanen reduction,
- Vening Meinesz reduction.

In addition to these reductions, also geological and other corrections are used.

10.3 Free-Air Reduction

This reduction takes into account only the height above sea level of the point of measurement. The purpose is to determine gravity acceleration g at sea level, if the corresponding value g_h is known at an altitude h; see Fig. 10.1.

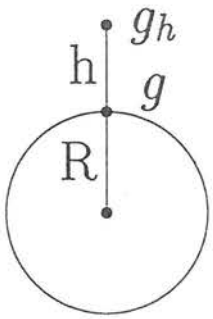

Figure 10.1: Calculation of the free-air reduction.

We shall consider the situation as if there were no masses between the sea level and the point of observation (only free air, the gravitational effect of which is negligible). To obtain just the main term of this reduction, it will be sufficient to consider the Earth as a sphere. Denoting M and R the mass and radius of the Earth's, respectively, we have

$$g = \frac{GM}{R^2} \,, \tag{10.1}$$

$$g_h = \frac{GM}{(R+h)^2} = \frac{GM}{R^2\left(1+h/R\right)^2} \doteq \frac{GM}{R^2}\left(1-2\frac{h}{R}\right) = g - \frac{2gh}{R} \,. \tag{10.2}$$

Consequently, for the gravity field at sea level we get

$$g = g_h + \delta_F \,, \tag{10.3}$$

where

$$\boxed{\delta_F = \frac{2g}{R}h} \tag{10.4}$$

is the *free-air reduction* (the Faye reduction).

Now substitute (10.1) into (10.4), and replace R by equatorial radius a. Using the numerical values for the Geodetic Reference System 1980, i.e. $GM = 398\,600.47 \times 10^9\ \mathrm{m^3 s^{-2}}$, $a = 6\,378\,137$ m, we arrive at

$$\delta_F = 3.072h\ ,$$

where h is given in metres and δ_F in $\mu\mathrm{m\,s^{-2}}$, or

$$\delta_F = 0.3072h\ ,$$

where h is again in metres and δ_F in milligals ($1\ \mathrm{mgal} = 10\ \mu\mathrm{m\,s^{-2}}$). The latter formulae do not take into account the rotation of the Earth and its flattening. By inserting mean values into (10.4), for example, $g = 9.81\ \mathrm{m\,s^{-2}}$ and $R = 6\,371$ km, we arrive at another formula:

$$\delta_F = 0.3080h\ ,$$

where h is in metres and δ_F in milligals.

A more accurate formula for the free-air reduction, used in some countries, is (Pick, 1987)

$$\boxed{\delta_F = 0.308\,55h - 0.000\,000\,072h^2 + 0.000\,219h\cos 2\phi}\ , \qquad (10.5)$$

where h is the height above sea level in metres, ϕ is the geographic latitude, and reduction δ_F is in milligals; see formula (7.57) in Section 7.5. The last two terms of the latter formula are often omitted (specially for small heights above sea level and ϕ around 45°). We then get a simpler formula,

$$\boxed{\delta_F = 0.3086h}\ , \qquad (10.6)$$

which is often referred to in the literature; see Heiskanen and Vening Meinesz (1958), Garland (1965) and others.

Suppose that the measured gravity g_h is reduced to sea level with the use of the free-air term only. The *free-air gravity anomaly* is then defined as the deviation of the reduced value from normal gravity γ, i.e.

$$\Delta g_F = g - \gamma = \left(g_h + \delta_F\right) - \gamma\ . \qquad (10.7)$$

Taking into account (10.6) we have

$$\boxed{\Delta g_F = \left(g_h + 0.3086h\right) - \gamma}\ . \qquad (10.8)$$

Free-air anomalies are used as input data in solving problems of geodetic gravimetry (determination of the figure of the Earth, determination of the heights above sea level, and others).

The maps of the free-air anomalies are, to some extent, similar to topographic maps. For example, free-air anomalies are generally positive in high mountains. This indicates that the effect of the masses above sea level must also be taken into account. This is important especially in gravimetric prospecting.

10.4 Bouguer Reduction without the Topographic Correction

As a first approximation of the form of the topographic masses, we shall consider only the so-called Bouguer plate (Fig. 10.2). It is a horizontal plate of thickness h, which is equal to the height above sea level of point P, where the reduction is being calculated. We assume that the plate is homogeneous, and has a constant density, ρ.

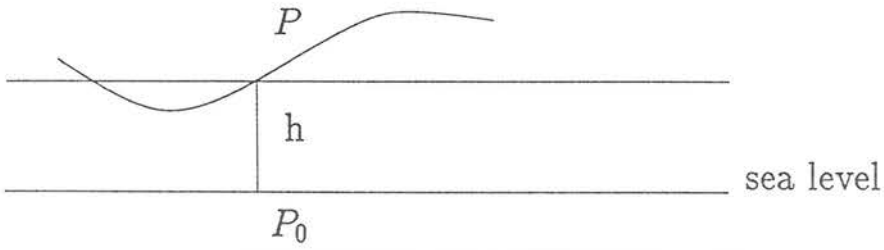

Figure 10.2: The Bouguer plate approximating the real terrain.

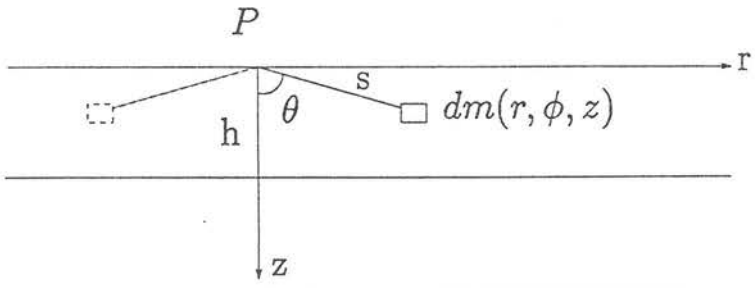

Figure 10.3: Calculation of the Bouguer term.

Let us calculate the gravitational effect of the plate (Fig. 10.3) The contribution to the intensity at point P due to mass element dm is of magnitude

$$\mathrm{d}g_t = \frac{G\,\mathrm{d}m}{s^2}\,, \tag{10.9}$$

and its direction is from P towards dm; $s = \sqrt{r^2 + z^2}$ is the distance of element dm from P. However, only the vertical component contributes to the resultant gravitational field due to the symmetry about the z-axis (to each element dm there is a symmetrical element, i.e. the dashed element in Fig. 10.3, eliminating the horizontal component at point P). Thus, the corresponding vertical component is

$$\mathrm{d}g = \mathrm{d}g_t \cos\theta = \mathrm{d}g_t \frac{z}{s} \,. \tag{10.10}$$

Putting $\mathrm{d}m = \rho \mathrm{d}V$, where $\mathrm{d}V$ is the corresponding volume element, we arrive at the total intensity at P due to the whole plate:

$$g = G\rho \iiint_V \frac{z}{s^3} \mathrm{d}V \,. \tag{10.11}$$

From the integral calculus we know that the volume element may be expressed in cylindrical coordinates as $\mathrm{d}V = r\,\mathrm{d}r\,\mathrm{d}\lambda\,\mathrm{d}z$. Hence, the gravitational field of the plate is

$$g = G\rho \int_0^{\infty} \int_0^{2\pi} \int_0^{h} \frac{zr}{\left(r^2 + z^2\right)^{3/2}} \mathrm{d}r\,\mathrm{d}\lambda\,\mathrm{d}z \,. \tag{10.12}$$

Performing first the integration over angle λ, then over r, and finally over z, we arrive at

$$g = 2\pi G\rho \int_0^{h} \left[\frac{-z}{\left(r^2 + z^2\right)^{1/2}} \right]_{r=0}^{\infty} \mathrm{d}z = 2\pi G\rho \int_0^{h} \mathrm{d}z = 2\pi G\rho h \,. \tag{10.13}$$

The attraction of the Bouguer plate increases the gravity at point P. Therefore, this value must be subtracted from the measured values. Consequently, the *Bouguer term* of the gravity reduction is

$$\boxed{\delta_B = -2\pi G\rho h} \,. \tag{10.14}$$

This term reduces the effect of the Faye term by about one-third,

$$\delta_F + \delta_B = \left(0.3086 - 0.0419\rho\right)h \,, \tag{10.15}$$

where ρ is in g cm^{-3}, h in metres and the gravity reduction in milligals. Again taking $\rho = 2.67$ g cm^{-3}, we get

$$\delta_F + \delta_B = (0.3086 - 0.1119)h = 0.1967h \ . \qquad (10.16)$$

This sum of the Faye and Bouguer terms, $\delta_F + \delta_B$, is called the *Bouguer reduction without the topographic correction*. The Bouguer anomaly without the topographic correction is then defined as

$$\boxed{(\Delta g)_B = g_h + \delta_F + \delta_B - \gamma} \ . \qquad (10.17)$$

This method of reducing the gravity measurements is used in interpretations on territories of a minor extent where terrain irregularities and the curvature of the Earth need not be taken into account. If the neighbourhood of point P is flat, the error due to terrain irregularities may by neglected. Otherwise, this Bouguer anomaly contains a certain negative error. In areas of hills or deep valleys, these errors produce spurious negative anomalies. Their influence can be removed by introducing terrain corrections; see the next section.

Before discussing the terrain corrections, let us calculate the effect of the Earth's curvature, and describe another method of deriving the Bouguer term (10.14), which is based on Gauss' law.

If the territory under consideration has a radius of more than about 80 km, the effect of the Earth's curvature should be taken into account. The plane Bouguer plate should then be replaced by a spherical layer of the same thickness h and constant density ρ. Take the mean radius of the Earth, R, as the inner radius of the layer (Fig. 10.4). Let us calculate the gravitational field of the spherical layer on the outer surface, i.e. at distance $(R+h)$ from the centre.

Contrary to Fig. 10.1, where the space between radii R and $(R+h)$ was empty, here we consider a constant positive density ρ. The mass of the layer is

$$m = \frac{4}{3}\pi\rho\left[(R+h)^3 - R^3\right] . \qquad (10.18)$$

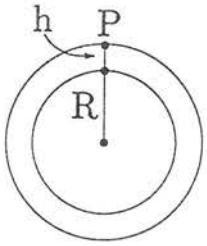

Figure 10.4: Spherical layer.

The intensity of the gravitational field due to this layer, at point P on its surface, is

$$g = \frac{Gm}{(R+h)^2} = \frac{4}{3}\pi G\rho\left[R + h - \frac{R^3}{(R+h)^2}\right] . \qquad (10.19)$$

The last term in this expression may be expressed as

$$\frac{R^3}{(R+h)^2} = R\left(1+\frac{h}{R}\right)^{-2} = R\left[1-2\frac{h}{R}+3\left(\frac{h}{R}\right)^3+\cdots\right].$$

By substituting it into (10.19), and changing the sign, we arrive at the Bouguer term for the spherical layer:

$$\delta_B^S = -4\pi G\rho h\left[1-\frac{h}{R}+\cdots\right]. \qquad (10.20)$$

The corresponding gravity anomaly would be given by formula (10.17) where δ_B would be replaced by δ_B^S. In practice, however, formula (10.14) for the flat layer is used most often.

Note a curious fact that the flat Bouguer layer leads to a reduction of $-2\pi G\rho h$, whereas the spherical layer leads to an anomaly, whose first term is twice larger, i.e. $-4\pi G\rho h$. This means that there is no continuous transition between these formulae. We shall return to this serious problem below in this chapter.

Let us return to formula (10.13) for the gravitational field of the flat Bouguer plate. We would like to show that this formula can easily be derived also with the aid of Gauss' law.

Consider a cylinder which is cut out from the Bouguer plate, its axis coinciding with the *z*-axis in Fig. 10.3, and its radius *r* being arbitrary (Fig. 10.5). According to Gauss' law, the flux of the intensity through the surface of the cylinder is

$$\iint_S \mathbf{E}\cdot\mathbf{dS} = -4\pi GM\ , \qquad (10.21)$$

where M is the mass inside the cylinder, and vector **dS** has the direction of the outward normal. Since the intensity is parallel to the *z*-axis, the flux through the lateral area of the cylinder is zero, and the only non-zero contributions to the integral in (10.21) come from the bases of the cylinder. Equation (10.21) may then be expressed as

$$2\pi r^2 E_n = -4\pi G\pi r^2 h\rho\ , \qquad (10.22)$$

where E_n is the component of the intensity in the direction of the outward normal. However, the gravitational acceleration due to the plate has the direction of the inward normal. After changing the sign, $g=-E_n$, we immediately arrive at formula (10.13).

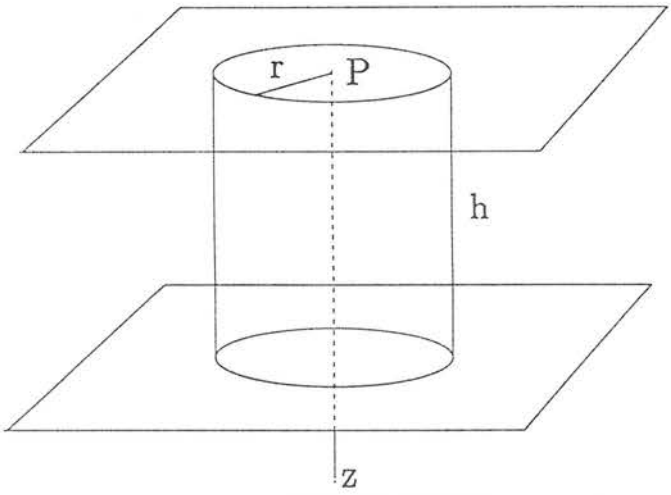

Figure 10.5: The cylinder to which Gauss' law is applied.

10.5 Topographic Correction. Complete Bouguer Reduction

The Bouguer term (10.14) only compensates the attraction of a layer. It is further necessary to introduce a correction for terrain irregularities (Fig. 10.6).

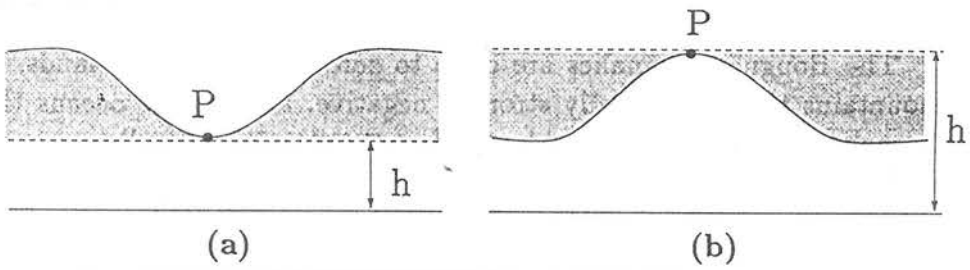

Figure 10.6: The masses considered in topographic corrections.

The presence of masses at higher altitudes than the altitude of the observer at point P reduces the gravity measured at P (Fig. 10.6a). To eliminate this effect we must add a positive topographic correction.

If the masses below height h are deficient (Fig. 10.6b), the result is similar. Again we should remove the gravitational effect of all masses above sea level. However, in the Bouguer term we have subtracted more than we should have done; we have subtracted the effect of masses which really do not exist. To correct this situation, we must add a positive term. Consequently, the topographic corrections are always positive.

The calculation of the topographic correction, δ_{top}, is usually performed by dividing the terrain into zones (rings), each zone into sectors, the mean height of each sector being determined from the map. The attraction of each sector is then

calculated by means of the known formulae. An example of this method is described in the next section.

In plains, topographic corrections may often be neglected, but not in high mountains where they may amount to several tens of milligals.

The sum of Faye reduction δ_F, Bouguer term δ_B and topographic correction δ_{top}, i.e.

$$\delta_F + \delta_B + \delta_{top} \ , \tag{10.23}$$

is called the complete Bouguer reduction, the Bouguer reduction with terrain correction, or only the Bouguer reduction. The *complete Bouguer anomaly* is then defined as

$$\boxed{(\Delta g_B) = g_h + \delta_F + \delta_B + \delta_{top} - \gamma} \ . \tag{10.24}$$

If we speak only of the Bouguer anomaly, we usually mean the complete Bouguer anomaly (10.24), i.e. the anomaly including the topographic term, not the simplified case of (10.17).

Bouguer anomalies are close to zero only in lowlands. In mountains they are usually highly negative, and on oceans they are highly positive. This is caused by the fact that the isostatically compensating masses are not considered in the Bouguer anomalies.

Bouguer anomalies are widely used in applied geophysics. However, serious physical objections may arise to their applications in geodesy; see the analyses at the end of this chapter.

10.6 An Example of Practical Computing of the Gravity Topographic Correction

To compute the gravity topographic correction correctly, it would be necessary to integrate over the whole surface of the Earth. However, for the purposes of exploration geophysics it is sufficient to integrate to quite a small distance. Sometimes the computation is carried out only to a distance of 20 or 30 km. We shall describe a more accurate method which considers the effects of topography up to a distance of 166.7 km. This method was developed at several institutes in the Czech Republic (Pick et al., 1973; Mareš et al., 1979).

The computation of the gravity topographic correction is still a very laborious problem. Means are therefore being sought to facilitate this process. In the method mentioned here, the neighbourhood of point P, at which the topographic correction is being determined, is divided into three parts. The topographic correction $\delta_{top} \equiv T$ is then the sum of the corresponding partial corrections, $T = T_1 + T_2 + T_3$. Term T_1 is the topographic correction in a square vicinity from 0 to 250 m, i.e. inside a square with a side of 500 m. Term T_2 is the correction for the region external to the square up to a distance of 5.24 km.

Term T_3 is the correction for the external region from 5.24 km to 166.7 km. The complete Bouguer anomaly is then defined in the Czech Republic by the formula

$$\Delta g = g_h - \gamma + (0.3086 - 0.0419\rho)h + \delta_{top} - B \ , \qquad (10.25)$$

where $B = B(h, \rho)$ is a correction for sphericity, referred to as Bullard's term. This term is the difference between the attraction of the segment of the spherical layer which extends to the distance of 166.7 km, and the attraction of the Bouguer plane (infinite) layer. Detailed tables of this term are available; see also Fig.10.7.

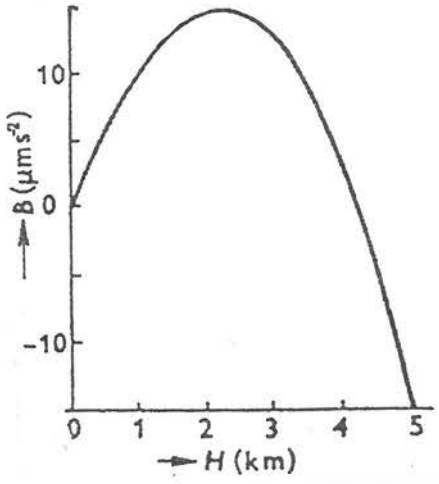

Figure 10.7: Bullard's term for $\rho = 2.67$ g cm^{-3} .

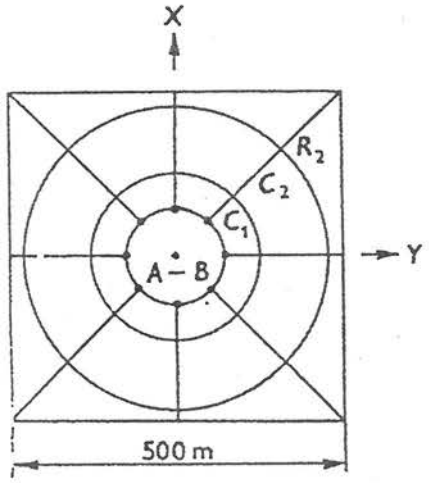

Figure 10.8: Zone T_1 and its division into sectors.

The internal correction T_1 forms a substantial part of the whole correction. Therefore, maximum care must be given to it. To compute it, the integration region (the square sides of 500 m) is divided into the sectors shown in Fig. 10.8.

Zone $A-B$ in the closest vicinity of point P is bounded by a radius of 68 m, zone C_1 lies between the radii of 68 and 130 m, zone C_2 between the radii of 130 and 230 m, and zone R fills up the remaining area of the square. The correction in respect of the central zone $A-B$ is determined by means of models. Three types of models are distinguished according to the morphology of the terrain in the vicinity of the gravity point: 1) regular slopes (two semi-circular wedges, characterized by two height differences Δh, describe the terrain); 2) irregular slopes (the height differences are determined at eight points); 3) convex and concave forms (again eight height differences are determined).

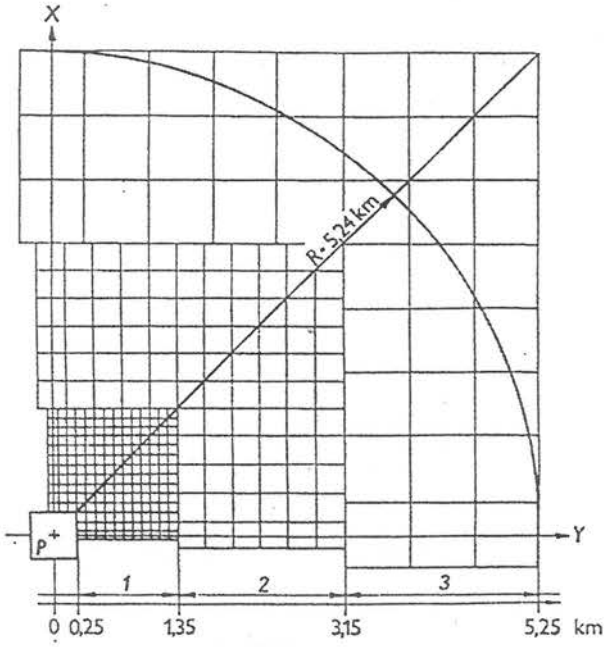

Figure 10.9: Location of the prisms for computing the terrain correction T_2 in the first quadrant: 1- prisms over 100 m; 2 - prisms over 300 m; 3 - prisms over 700 m.

Topographic correction T_2 is determined as the sum of the gravitational effects of elementary vertical prisms, which approximate the terrain shapes above and below the level of point P in the region which continues from the inner square up to a distance of 5.24 km (Fig. 10.9). Between the Cartesian coordinates of 0.25 and 1.35 km, the division is over 100 m, between 1.35 and 3.15 km over 300 m, and at greater distances over 700 m. Consequently, the mean heights above sea level for prisms with a base of 100×100 m must be determined for the whole territory as input data for such computations.

Correction T_3 for the range between 5.24 km and 166.7 km is expressed as a sum of two terms,

$$T_3 = f(h) + F(\phi, \lambda) ,$$

where $f(h)$ is the normal dependence of the regional correction on height above sea level h, and $F(\phi,\lambda)$ is the anomalous component of the correction dependent on geographic coordinates ϕ and λ. Special maps on a 1:200 000 scale were produced to determine $F(\phi,\lambda)$ for the territory under consideration; function $f(h)$ is also given for each sheet of these maps.

Note that the unusual distance of 166.7 km corresponds to one of Hayford's zones (Pick et al., 1973).

10.7 Isostatic Reductions

These reductions include the attraction of the mass above sea level, and the "attraction" of the mass deficiency up to the level of compensation. Traditionally, these reductions were calculated as contributions of cylindrical rings, which were further subdivided into segments.

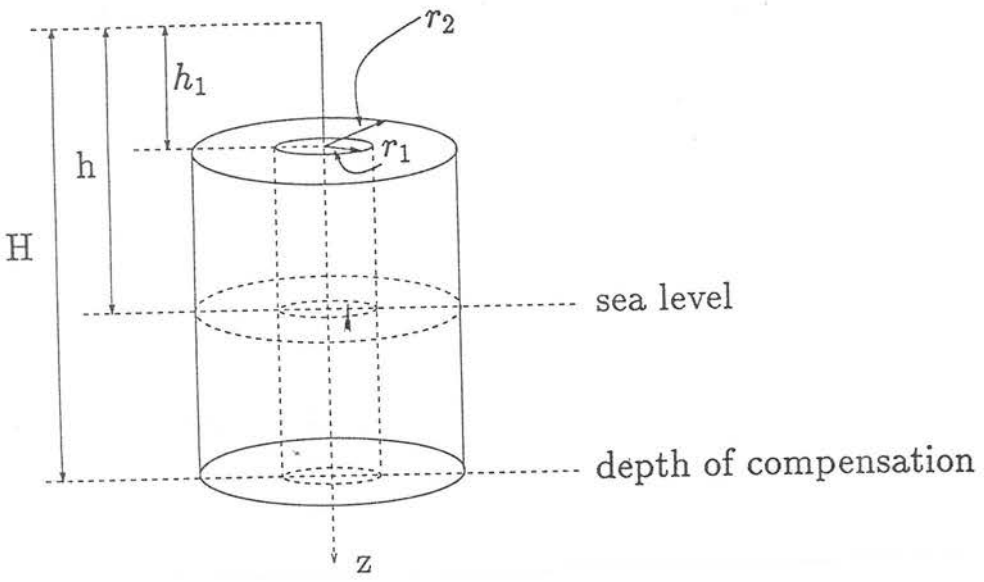

Figure 10.10: Cylindrical ring, for which the isostatic reduction is calculated.

Therefore, let us calculate the gravitational field of the cylindrical rings shown in Fig. 10.10. Let us choose the origin of the coordinate system at the point of observation P. Consider the cylindrical ring above sea level with density ρ. Calculate the intensity of its gravitational field, g_T, at point P (subscript T with g_T denotes topographic masses). The corresponding formula represents a generalized case of formula (10.12), i.e.

$$g_T = G\rho\int_{r_1}^{r_2}\int_{0}^{2\pi}\int_{h_1}^{h}\frac{zr}{\left(r^2+z^2\right)^{3/2}}\,\mathrm{d}r\,\mathrm{d}\lambda\,\mathrm{d}z = 2\pi G\rho\int_{h_1}^{h} z\,\mathrm{d}z\left[\frac{-1}{\left(r^2+z^2\right)^{1/2}}\right]_{r=r_1}^{r_2} =$$

$$= 2\pi G\rho\left(\sqrt{r_1^2+h^2}-\sqrt{r_1^2+h_1^2}-\sqrt{r_2^2+h^2}+\sqrt{r_2^2+h_1^2}\right). \tag{10.26}$$

The sum of the terms of this type yields the Bouguer and the topographic terms.

Isostasy requires another term to be added. Consider the Pratt-Hayford isostatic system. As a consequence of the compensation, the ring is not of density ρ, but along the whole depth from h_1 to H its density is changed by $\Delta\rho = -\dfrac{h-h_1}{H-h_1}\rho$. The "attraction" of this mass deficiency (of the compensating masses) is analogously

$$g_C = -2\pi G\frac{h-h_1}{H-h_1}\rho\left(\sqrt{r_1^2+H^2}-\sqrt{r_1^2+h_1^2}-\sqrt{r_2^2+H^2}+\sqrt{r_2^2+h_1^2}\right). \tag{10.27}$$

The isostatic reduction is then

$$\delta_{iso} = -\sum\left(g_T + g_C\right), \tag{10.28}$$

where the summation is carried out over all zones and sectors.

The isostatic reduction contains the Bouguer term and the topographic correction. Therefore, it is necessary to add only the Faye reduction. Consequently, the isostatic gravity anomaly is defined as

$$\left(\Delta g\right)_{iso} = g_h + \delta_F + \delta_{iso} - \gamma\ . \tag{10.29}$$

Various isostatic anomalies will be obtained for various values of compensation depth H. The isostatic model is often sought which explains the observed gravity data best, i.e. which yields the smallest gravity anomalies. For example, Hayford (1909) found that the sum of squares of the anomalies on the territory of the USA had a minimum for $H = 113.7$ km. This depth was, therefore, adopted as the best depth of compensation for the territory of the USA.

In computing the Airy-Heiskanen reduction, the same division into zones is used and the formulae are also analogous. There is a difference only in the mechanism of compensation. Depth H may be interpreted here as the thickness of the Earth's crust. Investigations have shown that the Airy-Heiskanen anomalies take minimum values for H between 20 and 40 km. These values may be used as first estimates of the crustal thickness.

An advantage of the isostatic anomalies is that they change only slowly from point to point. Consequently, one gravity measurement is sufficiently representative of a rather large area. These anomalies were used in determining the figure of the Earth (although their application occasionally led to deformations of the equipotential surfaces amounting to tens of metres), and also in determining the general properties of the Earth's crust on larger territories. However, we now know that there are too many deviations from the isostatic equilibrium in the mass distribution, so that isostatic anomalies are now used only exceptionally.

Some other reductions are not used now either, e.g. the Rudzki reduction, which transforms the masses above the Earth sphere into its interior by means of spherical inversion (Heiskanen and Vening Meinesz, 1958).

10.8 A Physical Analysis of the Gravity Reductions

Several problems appeared in the previous sections which should be analysed in greater detail, in particular:

1. To what extent do the reduced values represent the gravity field on the geoid ?
2. Why are the reductions for the plane Bouguer plate and the spherical layer so different ?
3. What is the relation between the individual anomalies ?

Let us try to answer these questions.

10.8.1 Two interpretations of gravity reductions

The primary purpose of introducing the reductions was to reduce the measured values to the geoid. However, in this procedure we have not used the real gradient of the gravity field, but a theoretical one for a concrete model. For example, for the simple case of the Faye anomaly we have used

$$(\Delta g)_F = (g_h + \delta_F) - \gamma \ . \tag{10.30}$$

The expression in the parentheses was interpreted as an approximation to the real gravity on the geoid.

However, the Faye reduction was a theoretical reduction. Consequently, we should rather combine it with the theoretical gravity γ, and put

$$(\Delta g)_F = g_h - (\gamma - \delta_F) \ ; \tag{10.31}$$

see Pick (1987). Value $\gamma - \delta_F$ may be interpreted as the normal gravity at height h above the spheroid (ellipsoid). Thus, in reducing the field, we actually do not reduce the real field downward to the geoid, but we rather extend the

theoretical field upward (by the same height h). This means that the reduced values should not be interpreted as the values on one equipotential surface.

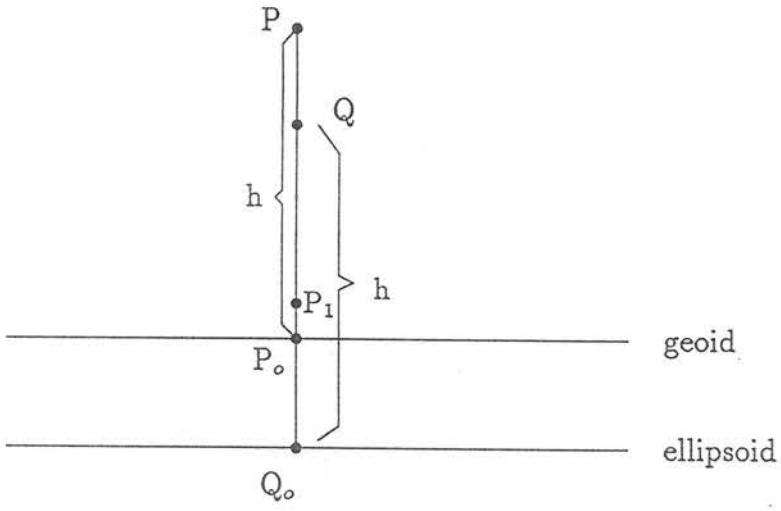

Figure 10.11: Various interpretations of gravity reductions.

The situation is shown in Fig. 10.11. Originally, the gravity anomaly should compare the real gravity on the geoid with the theoretical one (on the spheroid), i.e.

$$\Delta g = g(P_0) - \gamma(Q_0) \ . \tag{10.32}$$

However, the Faye anomaly (10.30) does not fulfil this requirement very well, because it uses the theoretical gradient instead of the real gradient. A better interpretation of this anomaly is given by formula (10.31), i.e.

$$(\Delta g)_F = g(P) - \gamma(Q) \ . \tag{10.33}$$

Nevertheless, the value $g_h + \delta_F$ in (10.30) may be considered as a reduction of the measured value $g(P)$, however not to point P_0 on the geoid, but to another point P_1 at distance h' from point P $\left(h' = \overline{P_1 P}\right)$. Namely, the term δ_F, which corresponds to distance h in the normal field, may be interpreted as a correction corresponding to another distance, h', in the real field:

$$\delta_F = -\frac{\partial \gamma}{\partial h} h = -\frac{\partial g}{\partial h} h' \ . \tag{10.34}$$

Note that in the first derivative we should put $\partial\gamma/\partial n$, $\mathbf{n}$ being the normal to the ellipsoid, but we neglect the deviations between the normals to the geoid and ellipsoid. The distance of point P_1 from the geoid, $h^* = \overline{P_0 P_1}$, comes out as

$$h^* = h - h' = \left[\left[\frac{\partial g}{\partial h} - \frac{\partial \gamma}{\partial h}\right] \Big/ \left(\frac{\partial g}{\partial h}\right)\right] h \,, \tag{10.35}$$

where h' has been expressed from (10.34). Since h^* depends on the height h of the point of measurement, gravity anomalies (10.30) are not, actually, related to one equipotential surface. This fact must be taken into account in some detailed calculations, e.g., in physical geodesy.

10.8.2 The Bouguer plate and spherical layer

Let us proceed to the second problem mentioned above. We should regard as inevitable from the physical point of view that the expression for a spherical problem transforms into the corresponding plane expression if the Earth's radius tends to infinity. However, this is not the case with formula (10.20) for the gravitational field of a spherical layer, which, in the limit for $R \to \infty$, yields a gravity field twice higher than formula (10.14) for the plane case.

In addition to this difference, there are other problems as well. The spherical layer, having a finite mass, increases the total mass of the Earth. However, by definition, the term containing the total mass of the Earth should be included in the normal gravity field. Therefore, the corresponding Bouguer reduction contradicts the normal gravity formula. In the case of the plane Bouguer plate, the situation is even worse, since we add a layer of infinite mass to a normal Earth.

We can therefore conclude that the Bouguer plate and Bouguer spherical layer represent two different models which may be useful in many applications (particularly in applied geophysics), but which are not compatible with the conception of the normal Earth. Thus, the Bouguer anomalies cannot be used in solving problems where the conception of the normal field is essential, especially in geodetic problems which require global data (Clairaut's theorem, Stokes' theorem, the Vening Meinesz formulae).

10.8.3 The Faye and Bouguer reductions as limiting cases of the isostatic reduction

Finally, we would like to show that the Faye and Bouguer reductions can be considered as two extreme cases of isostatic reductions. Consider formula (9.1) for the determination of the density in the Pratt-Hayford isostatic model, i.e.

$$\rho(H + h) = \rho_0 H \,. \tag{10.36}$$

For the corresponding density difference we had approximately

$$\Delta\rho = \rho - \rho_0 = -\frac{h}{H}\rho_0 \; ; \tag{10.37}$$

see formula (9.2).

In computing the Bouguer reductions we take the same density independently of the height above sea level, i.e. $\rho = \rho_0$, $\Delta\rho = 0$. The same result is obtained from (10.37) for $H \to \infty$. Thus, the Bouguer reduction is a special case of the isostatic reduction for the infinite compensation depth, $H = \infty$.

In computing the Faye reduction we consider the free air between the point of observation and sea level, i.e. $\rho = 0$. Inserting this value into (10.36) we get $H = 0$. Therefore, the Faye reduction may be regarded as an isostatic reduction with zero depth of compensation, $H = 0$.

This analysis also explains the opposite signs of the Faye and Bouguer anomalies in high mountains (the Faye anomalies are generally positive, whereas the Bouguer anomalies are generally negative). If the compensation depth is suitably chosen, the isostatic anomalies will be close to zero. However, in the case of the Faye anomalies the compensation depth is too small, so that we may expect a systematic shift of the numerical values of the anomalies. In the case of the Bouguer anomalies the compensation depth is too large and, therefore, the opposite sign of the anomalies should be expected.

10.9 Applications of Various Gravity Anomalies

The choice of an anomaly depends on the purpose, for which the anomalies are studied, and on the required accuracy. Anomalies are usually used for the following purposes:

a) determination of the flattening by means of Clairaut's theorem;
b) calculation of the undulations of the geoid by means of Stokes' theorem;
c) verification of hypotheses on isostasy;
d) investigation of anomalous masses within the Earth.

Let us discuss briefly the individual problems.

The derivation of Clairaut's theorem was based on the assumption that there are no masses outside the spheroid. Consequently, we must neglect the masses above the geoid and, therefore, we should use the Faye gravity reduction. We must then introduce another small reduction which will reduce the values on the geoid to the values on the spheroid.

There are various views as to the reduction to be used in Stokes' theorem. Some investigators propose only the Faye reduction, whereas others recommend the isostatic reductions. In both cases there are arguments for and against.

In verifying an isostatic hypothesis we shall use the corresponding isostatic reduction. We shall prefer the isostatic system which yields the smallest isostatic anomalies, i.e. explains the observed data best.

For studying anomalous masses in applied gravimetry, Bouguer anomalies are usually used.

Chapter 11

Interpretation of Gravity Anomalies

The initial data for the gravimetric interpretation are mostly maps of gravity anomalies. Maps of different scales are used for solving various geological problems. Gravity maps on scales of 1 : 50 000 to 1 : 200 000 with contour intervals of 1–2 mgal are used to study the subsurface geological structure down to a depth of several kilometres. Gravity maps on scales of 1 : 500 000 to 1 : 2 500 000 and smaller with contour intervals of 5–20 mgal are used to explore the structure of the Earth's crust as a whole.

Isostatic anomalies are often used for gravimetric investigations of large regional structures. Otherwise, Bouguer anomalies are mostly used (Pick et al., 1973; Mareš et al., 1979). As an illustration, a map of Bouguer anomalies in central Europe is shown in Fig. 11.1. One remarkable feature on this map should be pointed out, namely the presence of extensive negative anomalies in the mountain regions of the Alps and Carpathians. This is another example of the isostatic compensation, we discussed in the previous chapters.

11.1 General Features of Gravimetric Interpretations

The classical procedure of gravimetric interpretation usually consists of two stages:

1) *qualitative interpretation*;
2) *quantitative interpretation*.

By qualitative interpretation we understand the determination of connections between the anomalous gravity field and the geological structure. In particular, the number and type of perturbing bodies are determined from the shape of isoanomalies. For example, if the isoanomalies have a circular shape, the anomalous body may be a sphere or a vertical cylinder. If the shape of the isoanomalies is prolonged, the perturbing body can be approximated by a horizontal cylinder, prism, belt or similar bodies. In the stage of quantitative interpretation, numerical parameters of the perturbing bodies are determined, such as the coordinates of the mass centre, total mass, geometric shape or detailed density distribution.

In connection with quantitative interpretation, one frequently encounters in the literature the terms of *forward* and *inverse gravimetric problems*. The forward gravimetric problem is the computation of the gravity field (computation of the potential and its first or higher derivatives) for a given density model. On the other hand, the inverse gravimetric problem represents the determination of the parameters of a density model from observed gravity data.

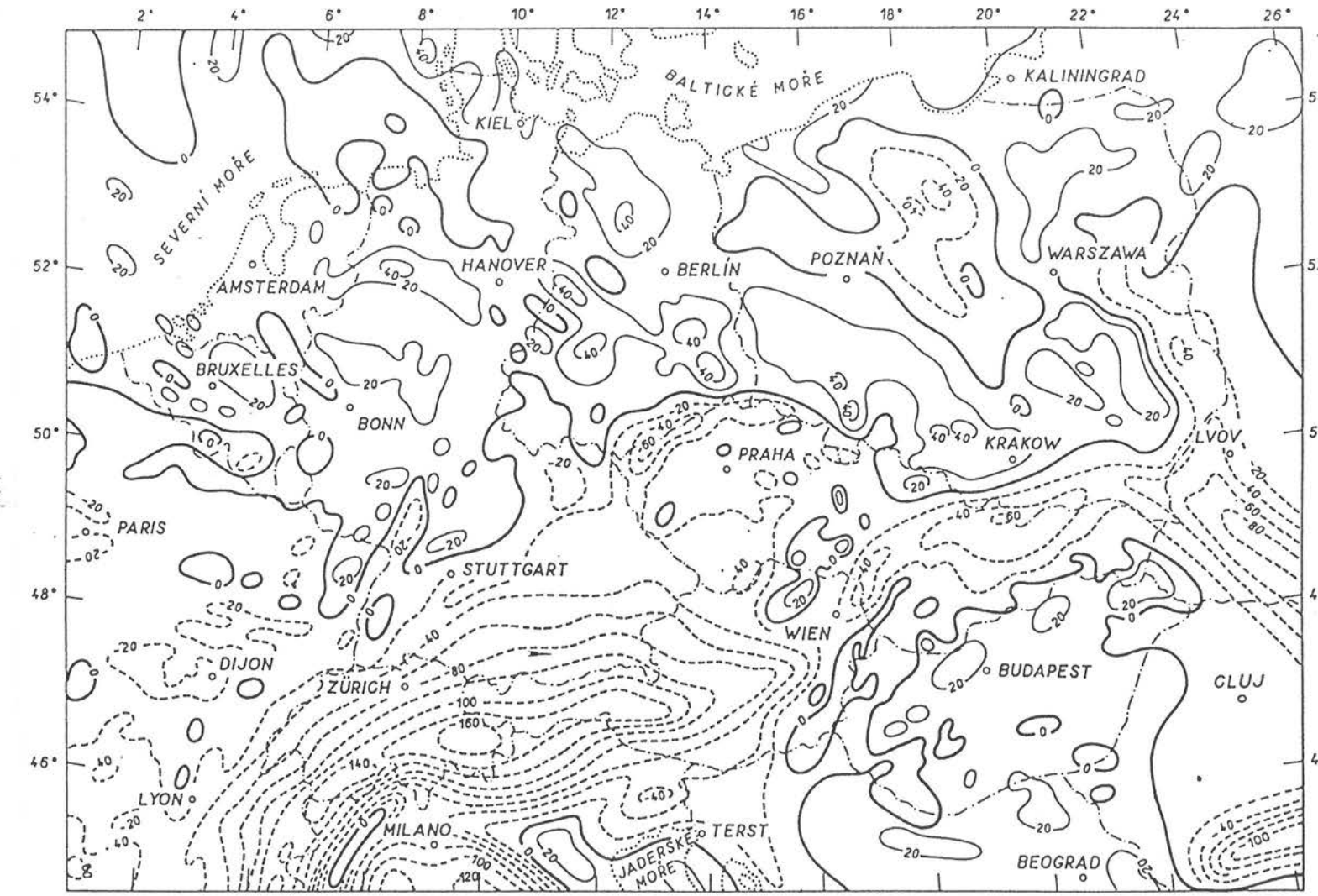

Figure 11.1: Gravimetric map of central Europe. The isolines of Bouguer anomalies are shown with the contours at intervals of 20 mgal. The solid lines show positive anomalies, and the dashed lines show negative anomalies. (After Ibrmajer (1978)).

The solutions of forward problems in gravimetry are relatively simple in comparison, for example, with complicated forward problems in seismology. Therefore, we could also expect the inverse gravimetric problems to be relatively simple. However, certain specific properties of the gravity field complicate the situation, in particular:

a) The gravitational field at any point is a superposition of the attractions due to nearby and distant masses.
b) Gravimetric interpretation is not unique.

Let us discuss some consequences of these properties.

Although gravitational attraction decreases with distance, even very distant bodies, if they are sufficiently massive, contribute significantly to the gravity field at any point of observation. The effects of nearby and distant bodies act simultaneously and their precise differentiation in the gravity field is impossible. Consequently, very large structural models, containing a large number of unknown parameters, should be considered in the process of gravimetric interpretation. It may even happen that the number of unknowns exceeds the number of measurements. Differential densities and parameters describing the dimensions and shape of the individual parts of the model may be considered as the unknowns. The mathematical formulation of the corresponding inverse problems then generally leads to large systems of equations which are usually ill-conditioned (the determinant of the system of linear equations is close to zero), or even underdetermined (the number of independent equations is less than the number of unknowns). To obtain a unique solution, we must either substantially simplify the model, which decreases its value for geological interpretations, or use also other data in the process of interpretation, such as geological data, measurements in boreholes, seismic and other measurements. To reduce these problems, we usually try to separate the effects of nearby and distant masses at least approximately by filtering the original gravity data; see the resolution of gravity anomalies, described below.

There are even deeper, principal causes of the ambiguity of gravimetric interpretations. We know, for example, that the gravitational field outside a homogeneous sphere is the same as if the whole mass of the sphere were concentrated into a point mass at the centre. The same is also true of a sphere which is inhomogeneous, but spherically symmetric (the density is only a function of the radial distance from the centre). Consequently, the gravity anomaly on the Earth's surface due to such a subsurface sphere does not contain any information about the size of the sphere. The interpretation of such an anomaly can yield only the coordinates of the centre of the sphere and its differential mass, but it cannot give its size or density; to determine these parameters, we need independent information from other geological or geophysical measurements. Let us now consider a fictitious sphere below the Earth's surface, and express the real density inside the sphere as a sum of the two following components: the first component is a spherically symmetric function; the second component is the corresponding difference. Move the masses of the first component in the radial direction so that a new spherically

symmetric distribution is obtained. This transport of masses will not change the external gravity field. In this way we can generate an infinite number of density models which all produce the same external gravity field. In interpreting the gravity data we are not able to distinguish between such models, which means that the inverse gravimetric problem is ambiguous.

As another example, consider the gravitational field of a plate. We have already seen that a homogeneous infinite layer (the Bouguer plate) produces a constant external gravitational field. Since the intensity of this field does not vary with distance from the layer, it is not possible to determine this distance from the gravity data. In other words, if such a layer contributes to the gravity field on the Earth's surface, we cannot determine its depth on the basis of gravity data only. If we change the depth of the layer, the gravity field on the Earth's surface will not change. This immediately indicates how to construct another category of gravimetrically equivalent models, and again demonstrates another type of ambiguity in the solution of the inverse gravimetric problem.

We have simplified the previous discussions a little by speaking only of unique and ambiguous formulations of inverse problems. In fact, the situation is more complicated. The uniqueness of the problems does not yet guarantee that its numerical solution is stable. A stronger condition must be fulfilled, namely the so-called *correctness* of the problem. It is said that a problem is posed correctly if its solution:

1) exists;
2) is unique;
3) is stable (continuously dependent on input data).

In our case, the stability means that a small change in the gravity anomaly (or in other initial quantities) would result only in a small change of the numerical solution of the inverse problem. It is evident that any ambiguous inverse problem is automatically also incorrect. However, even after simplifying the gravimetric problem and achieving a unique formulation, the problem may remain incorrect. The questions of solving incorrect problems have been discussed extensively in geophysics and mathematics in the last decades. Stable algorithms of solving inverse problems (e.g., various gradient methods) are now usually used, making it possible to solve even incorrect problems. However, another question arises, namely how to interpret the solution obtained, i.e. what is the relation of the computed model to the real structure.

It follows from the examples given above that we cannot determine the density distribution in vertically inhomogeneous models of the Earth only on the basis of surface gravity data. This is in sharp contrast with seismology where the construction of vertically inhomogeneous models of seismic velocities belongs to the basic tasks. These facts demonstrate the substantial limitations of gravimetric methods.

Since gravimetric interpretation cannot yield a complete density model of the medium on its own, supplementary data are needed. It is necessary to draw on all available geological and other geophysical data on the structure of the medium in the region of interest, and include them in the gravimetric

interpretation. For example, very detailed structural models can be derived from joint interpretations of seismic reflection data and gravity data. Profile reflection measurements can be used to determine the shape of subsurface interfaces, and the gravity data, keeping the shape of the interfaces fixed, can be used to determine the densities of individual geological bodies.

Among the main problems which gravimetry is helping to solve is the investigation of the shape of the basement of sedimentary basins and the course of tectonic faults and geological objects which differ in density from the surrounding medium. In investigating the deep structure of the Earth's crust, gravimetry deals with determining the shape of the Mohorovicic discontinuity and with solving problems connected with isostasy. Detailed structural models can be derived if the gravimetric methods are combined with seismic methods, as mentioned above.

The classical division of the interpretation process into qualitative and quantitative interpretations does not characterise the present situation well. Especially, the stage of qualitative interpretation, which is rather subjective and dependent on the experience of the interpreter, has gradually been replaced by many mathematical procedures which are more objective and can be realised by means of computers. Therefore, the process of contemporary gravimetric interpretations should rather be divided into the following steps:

1) *isolation*;
2) *identification*;
3) *interpretation*.

The first two steps represent processing gravity data before their quantitative interpretation, which is carried out in the third step. In the first step, by resolving the field of gravity anomalies, we try to isolate (separate) a valuable component from the measured data, and to suppress the other disturbing influences. In the stage of identification, we determine the number of perturbing bodies and some of their integral characteristics. Detailed numerical values are then determined in the last stage of interpretation. In the stages of the isolation and identification, various derived fields are computed from the measured gravity data. We shall describe the principles of these methods below. More detailed descriptions can be found, for example, in Pick et al. (1973), Mareš et al. (1979), and Parasnis (1997).

The interpretation process just mentioned is split into a sequence of partial transformations of the measured field, concluded by a quantitative interpretation. However, every transformation distorts the initial information and, consequently, also the final result. For this reason, there are also opposite tendencies, namely to interpret the original data without any transformations, even without gravity reductions. However, such formulations lead to very large systems of equations and to many mathematical problems. We shall not discuss these methods here, although we are convinced that in the future their importance will increase.

11.2 Resolution of Gravity Anomalies. Regional and Residual Anomalies

We have already discussed one method of resolving gravity fields, namely gravity reductions. In reducing gravity measurements, we separate the effects of the height above sea level, topographic masses, and also isostatically compensating masses.

We shall describe another approach which resolves the gravity anomaly, Δg, into a *regional anomaly*, Δg_R, and a *residual anomaly* (local anomaly), Δg_L:

$$\Delta g = \Delta g_R + \Delta g_L \ . \tag{11.1}$$

The regional anomaly should characterise the general trends of the field in the region under consideration. Therefore, the construction of the regional gravity map is a process which only shows the gravitational field due to the regional structure, i.e. a process used to remove local gravity effects. Consequently, the regional component of the field can be determined by smoothing and simplifying the original field. The computation of the regional and residual anomalies represents a certain type of filtration of the original field. The fields of the regional and residual anomalies are examples of derived fields.

Regional anomalies are usually used in solving problems of the deep geological structure. Residual anomalies, characterising details of the shallow structure, are used in gravimetric prospecting.

Note that there is no strict criterion for the resolution of the anomalous field into the regional and residual parts. This depends on our choice, but this choice should be adequate to the problem being solved. (A similar situation occurred when the resolution of the global gravity field into the normal and disturbing parts was also a question of agreement). Consequently, regional gravity may be quantitatively defined in a large number of ways, theoretically, in an infinite number of different ways (Griffin, 1949).

The determination of regional anomalies consists in smoothing the original anomalous gravity field. In the past, the smoothing was carried out graphically by smoothing and simplifying the isoanomalies in the gravity map directly. This method was fast, simple, and made it possible to take data on the geological structure into account. However, the result depended considerably on the subjective approach of the interpreter who was carrying out the smoothing. For this reason, the graphical methods were gradually replaced by numerical methods which guarantee a more "objective" solution of the problem.

Many numerical methods have been proposed for computing the regional anomalies. The "mean value" method and the method of reducing the gravity anomalies to a certain height above the Earth's surface are used most frequently. Let us describe a few of these methods.

11.2.1 The "mean value" method for a circle (Griffin's method)

The first precise definition of the regional anomaly was given by Griffin (1949). According to him, the regional anomaly at a given point P on a gravity map is defined as the averaged value of the gravity anomalies at a radial distance r from this point:

$$\Delta g_R(P) = \frac{1}{2\pi}\int_0^{2\pi} \Delta g(r,\alpha)\,\mathrm{d}\,\alpha \ , \tag{11.2}$$

where α is the corresponding angular coordinate. In other words, the regional anomaly at a given point is defined here as the mean value of the anomaly on a circle circumscribed about this point. Of course, the mean value (11.2) could have been defined on other figures than a circle, e.g. on rectangles, but the circle was used most frequently.

The choice of radius r controls the degree of smoothing; the smoothing generally increases with increasing value of r. Two limiting cases should be mentioned, namely $r = 0$ and $r \to \infty$. For $r = 0$, the averaged value becomes coincident with the initial value at point P, so that $\Delta g_R(P) = \Delta g(P)$, and the residual anomaly $\Delta g_L(P) = 0$. For r tending to infinity, and if the gravitational field under study is produced by a mass finite in size, the intensity decreases to zero, which yields $\Delta g_R(P) = 0$ and $\Delta g_L(P) = \Delta g(P)$.

Since, in general, the analytical form of $\Delta g(r,\alpha)$ is not known, some numerical method must be adopted for evaluating the integral in (11.2). Further, the numerical method chosen should be one economical of time. One such method, also proposed by Griffin, is to form the arithmetical average of a finite number of points on the circumference of the circle of radius r:

$$\Delta g_R(P) = \frac{1}{n}\left[\Delta g(r,\alpha_1) + \Delta g(r,\alpha_2) + \ldots + \Delta g(r,\alpha_n)\right] \ , \tag{11.3}$$

where n is the number of the points. This value is usually taken to be between 4 and 12.

Griffin analysed the properties of formula (11.3) on many maps of real anomalies. He used different radii of the circles, and 4, 6, 8 and 10 equidistant points on the given circle. Two Griffin's conclusions should be noted, since they were important for the further development of gravimetric methods, namely:

1. The value of the regional anomaly varies greatly with the radius of the circle.
2. The number of points used is only of secondary importance relative to the size of the circle.

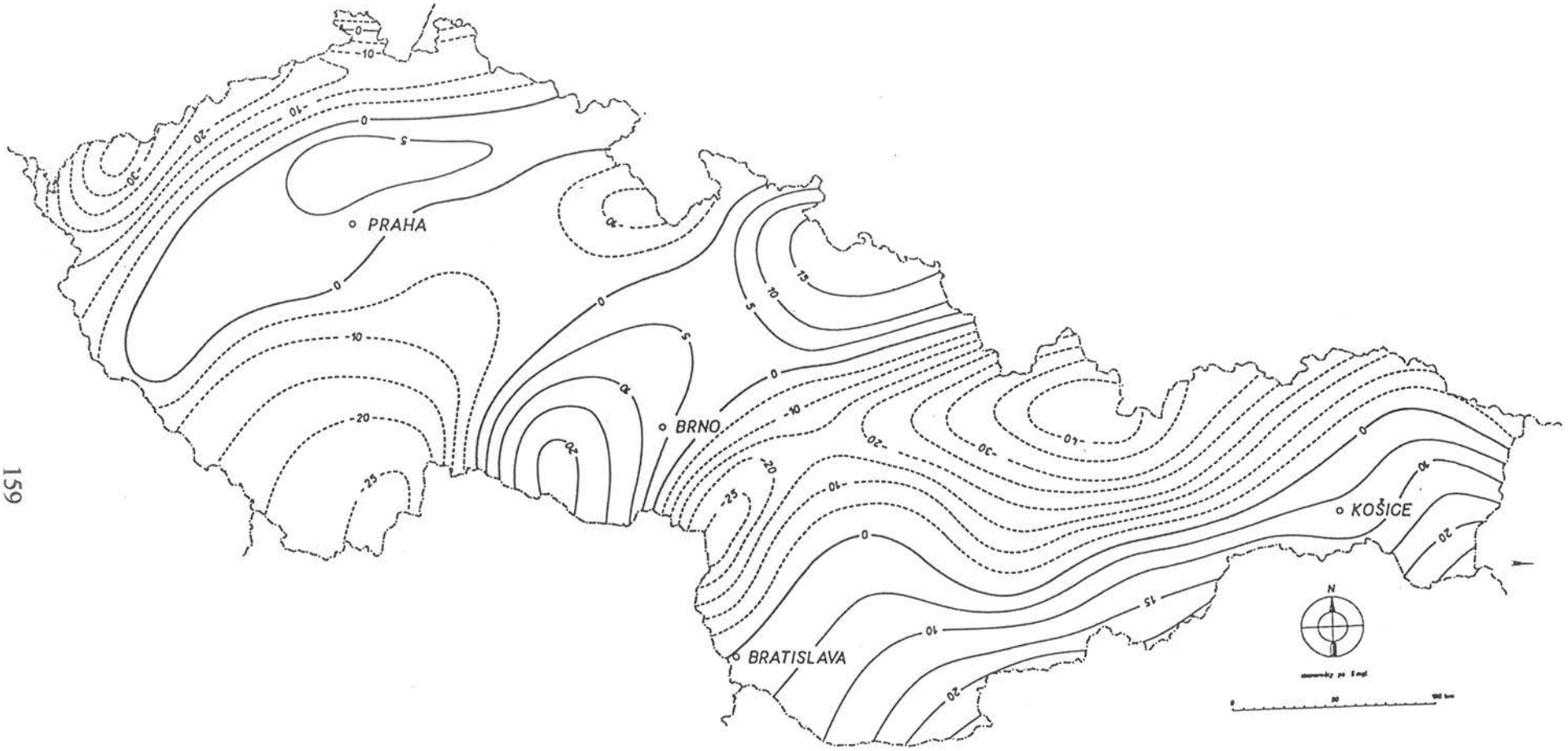

Figure 11.2: Map of the regional gravity anomalies on the territory of the former Czechoslovakia: Griffin's method, averaging along a circle $r = 8\sqrt{5}$ km. (After Ibrmajer (1978)).

Figure 11.3: Map of the residual gravity anomalies on the territory of the former Czechoslovakia; for details see Fig. 11.2.

The first of these properties raises the question whether the regional anomaly, computed in this way, is sufficiently objective, if its value depends greatly on a parameter of the data processing, r. For this reason, also other definitions of the regional anomaly were proposed and tested. As a result, the line integrals in the definition of the regional anomaly were replaced by areal integrals, the values of which depend less on the size of the integration domain. Some later textbooks, e.g., Mareš et al. (1979), do not even mention the older definitions of the regional anomalies by means of line integrals.

On the other hand, the second property, mentioned above, represents a positive finding. It indicates that even a slightly irregular distribution of the points along the circle could be acceptable. Consequently, many authors have used, for example, eight points as shown in Fig. 11.4b on a circle of radius $s\sqrt{5}$, where s is the step of a square grid. These points are not equidistant, but lie at the grid points of a rectangular grid, which has many advantages from the computational point of view. This method was also applied in computing the maps in Figs. 11.2 and 11.3, which we present here as an illustration.

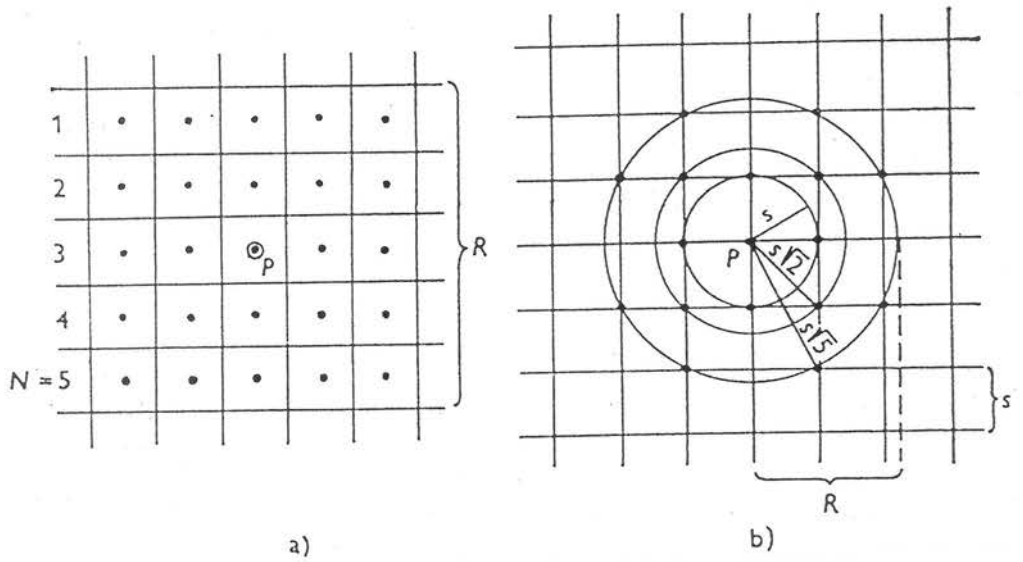

Figure 11.4: Integration domains for computing regional anomalies: a) a square of side R; b) the surface of a circle of radius R. (After Mareš et al. (1979)).

11.2.2 The "mean value" method for a square

The regional anomaly at point P can generally be defined as a weighted mean value of the field within a certain vicinity of the point:

$$\Delta g_R(P) = \frac{1}{\mu^*} \iint_S \mu \Delta g \, \mathrm{d}S \ , \tag{11.4}$$

where S is a surface vicinity of point P, μ is a weight function, and

$$\mu^* = \iint_S \mu \, \mathrm{d}S \ . \tag{11.5}$$

For the special case of $\mu = 1$ we get $\mu^* = S$; we shall restrict ourselves only to this case.

Consider the special case when region S is a square of side R; see Fig. 11.4a. Subdivide the square into N^2 partial squares of sides $s = R/N$. Determine the mean anomaly in each square, $(\Delta g)_{ij}$. Then the regional anomaly at the centre point P can be defined as the mean value in the whole square:

$$\Delta g_R(P) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\Delta g)_{ij} \ . \tag{11.6}$$

11.2.3 The "mean value" method for the surface of a circle

Calculate the regional anomaly by averaging over the surface of a circle of radius R. Using polar coordinates (and putting $\mu = 1$), the general formula (11.4) takes the form

$$\Delta g_R(P) = \frac{1}{\pi R^2} \int_0^R \int_0^{2\pi} \Delta g(r, \alpha) \, r \, \mathrm{d}r \, \mathrm{d}\alpha \ . \tag{11.7}$$

We shall express the latter formula by means of the mean values along concentric circles, see (11.2). For this purpose, denote expression (11.2) simply by $\overline{\Delta g}(r)$, i.e.

$$\overline{\Delta g}(r) = \frac{1}{2\pi} \int_0^{2\pi} \Delta g(r, \alpha) \, \mathrm{d}\alpha \ . \tag{11.8}$$

Formula (11.7) can then be expressed as

$$\Delta g_R(P) = \frac{2}{R^2} \int_0^R \overline{\Delta g}(r) \, r \, \mathrm{d}r \ . \tag{11.9}$$

For the numerical computation, divide the surface of the circle into n annuli by circles of radii $r_0 = 0, r_1, \ldots, r_n = R$. An example of three annuli is shown in Fig. 11.4b. Thus the i-th annulus is bounded by radii r_{i-1} and r_i. Further, denote the mean values (11.8) along the individual circles simply by

$$\overline{\Delta g}_i = \overline{\Delta g}(r_i), \quad i = 0, 1, \ldots, n \ . \tag{11.10}$$

As the mean value in the whole i-th annulus, many authors take the mean value of the anomalies along the inner and outer circles, i.e. the value

$$\frac{1}{2}\left(\overline{\Delta g}_{i-1} + \overline{\Delta g}_i\right) . \tag{11.11}$$

However, this formula may represent a rather rough approximation in many cases, in particular if the inner and outer radii are very different. (This problem will be critically discussed below). Formula (11.9) then becomes

$$\Delta g_R(P) = \frac{1}{R^2}\sum_{i=1}^{n}\left(\overline{\Delta g}_{i-1} + \overline{\Delta g}_i\right)\int_{r_{i-1}}^{r_i} r\,\mathrm{d}r = \frac{1}{2R^2}\sum_{i=1}^{n}\left(\overline{\Delta g}_{i-1} + \overline{\Delta g}_i\right)\left(r_i^2 - r_{i-1}^2\right) . \tag{11.12}$$

Write this expression in the form of two sums and introduce a new summation index $j = i - 1$ in the first sum:

$$\Delta g_R(P) = \frac{1}{2R^2}\left[\sum_{j=0}^{n-1}\overline{\Delta g}_j\left(r_{j+1}^2 - r_j^2\right) + \sum_{i=1}^{n}\overline{\Delta g}_i\left(r_i^2 - r_{i-1}^2\right)\right] . \tag{11.13}$$

In the first sum, again put i instead of j. In order to extend the summation in both terms from 0 to n, define formally the following radii:

$$r_{n+1} = r_n = R, \quad r_{-1} = r_0 = 0 \ . \tag{11.14}$$

We can then add the n-th term in the first sum of (11.13) and the zero-th term in the second sum, both being equal to zero. This yields

$$\Delta g_R(P) = \frac{1}{2R^2}\sum_{i=0}^{n}\overline{\Delta g}_i\left(r_{i+1}^2 - r_{i-1}^2\right) = \sum_{i=0}^{n} k_i\,\overline{\Delta g}_i \ , \tag{11.15}$$

where

$$k_i = \frac{r_{i+1}^2 - r_{i-1}^2}{2R^2} . \tag{11.16}$$

The numbers k_i have the character of weight coefficients. One can easily check that $\sum_{i=0}^{n} k_i = 1$.

As a concrete important case, let us consider the situation shown in Fig. 11.4b. The surface of the circle of radius R is divided there into three annuli by three circles which pass through the grid points of a square grid. The radii of the circles and the number of points in them, where the gravity anomalies are being considered, are as follows:

$r_0 = 0$, 1 point (centre point),

$r_1 = s$, 4 points,

$r_2 = s\sqrt{2}$, 4 points,

$r_3 = R = s\sqrt{5}$, 8 points.

Inserting these values into (11.14) to (11.16), we arrive at the final formula (Mareš et al., 1979)

$$\Delta g_R(P) = \frac{1}{10}\left(\overline{\Delta g}_0 + 2\overline{\Delta g}_1 + 4\overline{\Delta g}_2 + 3\overline{\Delta g}_3\right) , \qquad (11.17)$$

where $\overline{\Delta g}_0$ is the anomaly at the centre point P, $\overline{\Delta g}_0 = \Delta g(P)$. The anomalies at the grid points, mentioned above, are used to compute the mean values along the circles.

It is usually recommended that radius R be equal to two or three times the depths of the anomalous masses, the effect of which one wants to stress by means of the residual anomalies (Pick et al., 1973). Sometimes maps of the regional and residual anomalies are made for several different values of radius R, in order to find the most suitable map for the given problem.

Let us add several critical notes to formula (11.17). This formula was recommended for computing regional anomalies, for example, by Mareš et al. (1979). Many methods of computing "second derivatives" of the perturbing potential are based on this formula as well; see the following section and the review in Pick et al. (1973). The textbooks just mentioned do not contain a derivation of formula (11.17). We have derived this formula here in order to obtain a deeper insight into its properties. In particular, we would like to point out the unusual weighting of the anomalies and certain azimuth dependence.

At first sight, one property of formula (11.17) seems to be rather unusual, namely the coefficients with terms $\overline{\Delta g}_1$ and $\overline{\Delta g}_2$. Although radii r_1 and r_2 are relatively close to each other, the weight coefficient of the mean anomaly $\overline{\Delta g}_1$ is 2, whereas the coefficient of the more distant anomaly $\overline{\Delta g}_2$ is 4. We should rather expect the opposite. This property follows from formula (11.16) for determining k_i, which does not contain radius r_i, but only the squares of radii r_{i+1} and r_{i-1}. In our case, the large value of r_3^2 and the relatively small value of r_1^2 yield the large value of k_2. This surprising result is the consequence of the simplifying assumption used in (11.11), where the mean value of the gravity anomaly in each annulus is approximated by the arithmetic mean along the inner

and outer circles. This approximation may be substantiated only if the individual annuli are narrow (their outer and inner radii are not very different). However, this is not the case of the annuli in Fig. 11.4b, where the first and third annuli are rather wide.

The four points at distance r_1 in Fig. 11.4b are located on the grid lines passing through the centre point P, whereas the points at distance r_2 are located in the diagonal directions. Consequently, since coefficient k_1 is less than k_2, formula (11.17) underestimates the structures parallel to the grid lines and emphasises the structures in the diagonal directions. This azimuth filtering represents a drawback of formula (11.17). All these problems indicate that the approximation based on formula (11.11) should be critically analysed and other formulae should also be tested.

For example, the reader may object that the circle of radius $2s$ in Fig. 11.4b should also have been included. In this case, in particular, the coverage of the vicinity of point P would be more regular, since all grid points in the vicinity $r \le R$ would be considered. The radii of the circles and the number of points in them would then be as follows:

$r_0 = 0$, 1 point (centre point),
$r_1 = s$, 4 points,
$r_2 = s\sqrt{2}$, 4 points,
$r_3 = 2s$, 4 points,
$r_4 = R = s\sqrt{5}$, 8 points.

Formulae (11.15) and (11.16) now yield the formula for the regional anomaly in the form

$$\Delta g_R(P) = \frac{1}{10}\left(\overline{\Delta g}_0 + 2\overline{\Delta g}_1 + 3\overline{\Delta g}_2 + 3\overline{\Delta g}_3 + \overline{\Delta g}_4\right). \qquad (11.18)$$

In comparison with (11.17), the weight of the diagonally situated points has now been reduced substantially. For example, the mean anomaly in the most distant circle $r = R$ now has a weight of only 1/10, as opposed to 3/10 in formula (11.17). This analysis indicates that formula (11.18) may have better properties than the recommended formula (11.17). However, only tests of theoretical and real data may answer this question.

11.2.4 Polynomial and other approximations

The regional field in the area under consideration may be defined by a polynomial of the N-th degree as follows:

$$\Delta g_R(x, y) = \sum_{i=0}^{N} \sum_{j=0}^{N-i} a_{ij} x^i y^j \quad . \qquad (11.19)$$

For example, for $N = 2$ we have

$$\Delta g_R(x, y) = a_{00} + a_{01}y + a_{02}y^2 + a_{10}x + a_{11}xy + a_{20}x^2 \ . \qquad (11.20)$$

Coefficients a_{ij} can be determined, for example, by least-squares adjustment of the initial gravity anomalies.

The polynomial approximation of gravity anomalies represents a rather formal method, which does not take the specific properties of the gravity field into account. Consequently, some trends in the gravity anomalies cannot be satisfactorily approximated by a polynomial, no matter of what degree. An example is the anomaly produced by an infinitely long vertical step. More complicated functions, for example $\Delta g_R = C\tanh(mx)$, may be necessary in treating such trends analytically (Parasnis, 1997).

Instead of the double power series (11.19), series of orthogonal polynomials have also been proposed. Nevertheless, the computation of regional anomalies by means of polynomial approximations has not been widely used in practice.

Still more sophisticated procedures are based on the two-dimensional Fourier transformation. The regional anomaly may then be defined as the field of low-degree Fourier components. Nevertheless, even these mathematically elegant methods are rather formal. They are based on general mathematical transformations, which do not take the specific features of the gravity field into account.

From this point of view, the method described below could be classified as more "physical".

11.2.5 Upward analytic continuation

With increasing distance from a field source (anomalous mass), the amplitudes of gravity anomalies decrease, and the anomalous gravity field becomes smoothed. This property can be utilised to define the regional field Δg_R as the anomalous field at a certain height h above the Earth's surface (Fig. 11.5).

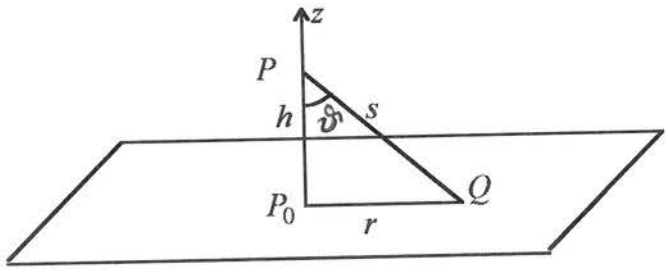

Figure 11.5: Geometry of the upward analytic continuation.

If the gravity anomaly is recalculated from a given level to a higher level (upward continuation), the derived field retains only the effects of more

extended structures, i.e. usually deeper structures. On the other hand, the recalculation of the anomalies to a lower level (downward continuation) stresses the effects of smaller and shallower bodies. In this connection we sometimes speak of "gravimetric focusing", since this procedure resembles the properties of an optic lens, which can also be focused onto distant or close objects.

Consider the Earth's surface to be a plane, and introduce a Cartesian coordinate system whose (x, y)-plane coincides with the Earth's surface, its origin being denoted by P_0, and the z-axis being oriented upward. Denote the gravity anomalies on the Earth's surface by $\Delta g(x, y, 0)$. The values of the anomalies in the upper half-space can be computed by means of Poisson's integral. For point P on the z-axis, this integral has the following special form:

$$\Delta g(P) = \Delta g(0,0,h) = \frac{h}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\Delta g(x,y,0)}{s^3} \mathrm{d}x\,\mathrm{d}y, \tag{11.21}$$

where $s = \sqrt{x^2 + y^2 + h^2}$ is the distance of point P from the corresponding point Q on the Earth's surface where the anomaly $\Delta g(x, y, 0)$ is being considered. An approximate derivation of formula (11.21) will be given below.

Poisson's integral (11.21) can easily be expressed in cylindrical coordinates r, α, z as follows (Pick et al., 1973):

$$\Delta g(0, h) = \frac{h}{2\pi} \int_0^{\infty} \int_0^{2\pi} \frac{\Delta g(r, \alpha, 0)}{\left(r^2 + h^2\right)^{3/2}} r\,\mathrm{d}r\,\mathrm{d}\alpha\ , \tag{11.22}$$

$\Delta g(0, h)$ denoting the anomaly at point P, where $r = 0$ and $z = h$. If the symbol $\overline{\Delta g}(r)$ is used to denote the mean value of the anomalies $\Delta g(r, \alpha, 0)$ along a circle of radius r, the following formula is obtained for computing the regional anomalies:

$$\Delta g(0, h) = h \int_0^{\infty} \frac{\overline{\Delta g}(r)}{\left(r^2 + h^2\right)^{3/2}} r\,\mathrm{d}r\ . \tag{11.23}$$

For the numerical computation we can use procedure analogous to that described above between formulae (11.10) and (11.16). Divide the integration interval into intervals $\Delta r_i = r_i - r_{i-1}$. Put $\overline{\Delta g}_i = \overline{\Delta g}(r_i)$. Approximate the mean value of the anomaly in the whole i-th annulus by the mean value along the inner and outer circles, i.e. by $\frac{1}{2}\left(\overline{\Delta g}_{i-1} + \overline{\Delta g}_i\right)$. We then obtain approximately

$$\Delta g(0,h)=\frac{h}{2}\sum_{i=1}^{\infty}\left(\overline{\Delta g}_{i-1}+\overline{\Delta g}_i\right)\int_{r_{i-1}}^{r_i}\frac{r\,\mathrm{d}r}{\left(r^2+h^2\right)^{3/2}}=$$

$$=\frac{h}{2}\sum_{i=1}^{\infty}\left(\overline{\Delta g}_{i-1}+\overline{\Delta g}_i\right)\left[\left(r_{i-1}^2+h^2\right)^{-1/2}-\left(r_i^2+h^2\right)^{-1/2}\right]=\sum_{i=0}^{\infty}K_i\overline{\Delta g}_i\,, \tag{11.24}$$

where

$$K_0=\frac{1}{2}\left[1-\frac{h}{\sqrt{r_1^2+h^2}}\right],\quad K_i=\frac{1}{2}\left[\frac{h}{\sqrt{r_{i-1}^2+h^2}}-\frac{h}{\sqrt{r_{i+1}^2+h^2}}\right]\text{ for } i>1.$$

For practical purposes, only a finite number of terms in the series is considered (Pick et al., 1973).

Special methods of the upward continuation for two-dimensional cases are mentioned in Mareš et al. (1979). For example, Andreev's simple method for the two-dimensional case uses the formula

$$\Delta g(0,h)=\frac{1}{n}\sum_{j=1}^{n}\overline{\Delta g}_j\,, \tag{11.25}$$

where $\overline{\Delta g}_j$ is the mean anomaly along the i-th segment of the profile, as seen from point P inside angle $\Delta\varphi=\frac{\pi}{n}$.

Let us return to Poisson's integral (11.21), describe the main steps of its derivation, and discuss some properties of the corresponding formulae.

In advanced chapters of potential theory, we can learn that the external gravitational potential of a bounded body (a volume potential) can be expressed as the sum of two surface potentials, namely the potentials of a single and double layer which are spread over the surface of the body. This property follows from Green's second theorem (Pick et al., 1973). If the surface encompassing the body is equipotential, a single layer is sufficient. Therefore, we shall try to construct such a single layer on the Earth's surface which produces the given gravity anomaly $\Delta g(x,y,0)$. This surface distribution, according to the above-mentioned theorem, produces the same external field as the masses below the Earth's surface.

We have seen that a homogeneous Bouguer plate produces a homogeneous field, the intensity of which is $E=2\pi G\rho H$, ρ being the density and H the thickness of the layer. Consider an inhomogeneous plate whose density varies only in the horizontal directions $\rho=\rho(x,y)$. Its field is more complicated than the field of the homogeneous plate. Compress the plate, i.e. decrease its

thickness H and increase density $\rho(x, y)$ in such a way that their product $\sigma(x, y) = H\rho(x, y)$ remains the same function of x and y. In the limit for H tending to zero we obtain a single layer with surface density $\sigma(x, y)$. At point $Q(x, y, 0)$ on the surface of the plate, the main contribution to the vertical component of intensity comes from the attraction of nearby masses since the attraction of distant masses is nearly horizontal. The field at point Q on the surface will be the same as if the single layer were homogeneous and had the same local density $\sigma(Q)$. Hence, the field on the layer is

$$E(x, y, 0) = 2\pi G\sigma(x, y) \,. \tag{11.26}$$

By substituting $\Delta g(x, y, 0)$ for $E(x, y, 0)$ we get

$$\sigma(x, y) = \frac{\Delta g(x, y, 0)}{2\pi G} \,. \tag{11.27}$$

Thus the given gravity anomaly $\Delta g(x, y, 0)$ can be considered as the field produced by a single layer of density (11.27). The vertical component of the intensity at height h is then (Fig. 11.5)

$$\Delta g(0, 0, h) = G \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\sigma(x, y)}{s^2} \cos\vartheta \, \mathrm{d}x \, \mathrm{d}y \,, \tag{11.28}$$

where ϑ is the angle between the vertical and direction PQ, $\cos\vartheta = h/s$. By inserting (11.27) into (11.28) we then arrive at Poisson's integral (11.21).

Consider the special case of $\Delta g(x, y, 0) = c$, where c is a constant. Formula (11.23) then yields

$$\Delta g(0, h) = hc \int_0^{\infty} \frac{r \, \mathrm{d}r}{\left(r^2 + h^2\right)^{3/2}} = hc \left[\frac{-1}{\sqrt{r^2 + h^2}} \right]_{r=0}^{\infty} = c \,. \tag{11.29}$$

This field does not change with height h, so that it is identical with the field of the Bouguer plate. This property does not correspond to the real field, which decreases with height, the decrease being described approximately by the Faye term. Consequently, Poisson's integral (11.21) does not describe the real field at height h, since at least the Faye term is missing there. Poisson's integral would describe the real field if the Earth were an infinite plate. In other words, although we speak of upward continuation of the anomalous field, we cannot interpret the result as the real field at height h. The transformation is more

formal. Parameter h cannot be interpreted as the height to which we reduce the real field from the Earth's surface, but rather as a distance from which we "see" the field on the Earth's surface. This substantiates the idea of "gravimetric focusing" for this procedure. On the other hand, it simplifies our situation, since we can define the regional anomaly directly by Poisson's integral (11.21), i.e. put

$$\Delta g_R(P_0) = \Delta g(P) \ , \tag{11.30}$$

where $\Delta g(P)$ is given by (11.21); see Fig. 11.5. Namely, if we knew the real anomalous field at height h, we could also use it to define the regional anomaly on the Earth's surface, but a reduction of Faye's type would be necessary.

11.3 Identification of Perturbing Bodies

As mentioned above, the purpose of this stage of the gravimetric interpretation is to obtain some basic parameters of the perturbing body or bodies, such as the number of perturbing bodies, their approximate shape, position of the centre of mass, and the total differential mass.

11.3.1 Downward analytic continuation

The upward analytic continuation is a process of smoothing of the anomalous field by transforming the field to a larger distance from its sources. The downward analytic continuation is the opposite process whereby we try to approach the sources and, consequently, to increase the contrasts in the original field. The places of a strong transformed field, or even the singularities of this field, then indicate the probable position of the perturbing bodies.

The downward analytic continuation is only an approximate transformation, since it is based on Laplace's equation, $\Delta V = 0$, whereas the field below the Earth's surface satisfies Poisson's equation, $\Delta V = -4\pi G\rho$.

Formula (11.21) for the upward analytic continuation relates the anomalies at levels $z = h$ and $z = 0$. The formula for the downward continuation should relate anomalies at levels $z = 0$ and $z = -h$. It is evident that a formula for the downward continuation can be obtained from (11.21) by reducing the z-coordinates by h:

$$\Delta g(0, 0, 0) = \frac{h}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\Delta g(x, y, -h)}{s^3} \mathrm{d}x\,\mathrm{d}y \ .$$

By changing the notation, this formula can be expressed in a more general form as

$$\Delta g(x, y, 0) = \frac{h}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\Delta g(\xi, \eta, -h)}{\left[(\xi - x)^2 + (\eta - y)^2 + h^2\right]^{3/2}} \mathrm{d}\xi \mathrm{d}\eta \; ; \qquad (11.31)$$

see Fig. 11.5. This is the formula required for the downward analytic continuation.

Despite the formal equivalence of formulae (11.21) and (11.31), there is a principal difference in their mathematical character. Namely, the given anomaly on the Earth's surface, $\Delta g(x, y, 0)$, stands in the integrand of (11.21), and we want to determine $\Delta g(0, 0, h)$. As opposed to this, the given $\Delta g(x, y, 0)$ is on the left-hand side of Eq. (11.31), but the unknown anomaly, $\Delta g(\xi, \eta, -h)$, appears in the integrand. In other words, the upward analytic continuation (11.21) represents an evaluation of a double integral, whereas the downward analytic continuation (11.31) represents a more complicated mathematical problem, namely the solution of an integral equation. We shall not deal with this difficult problem here, but we refer the reader, e.g., to Pick et al. (1973).

We shall describe a simpler method of the downward continuation for a two-dimensional case. The method is based on the mean value theorem for harmonic functions.

At the centre of a vertical circle C which is perpendicular to a two-dimensional field and has radius r, the anomaly can be expressed as

$$\Delta g(0, 0) = \frac{1}{2\pi} \int_C \Delta g(r, \alpha) \mathrm{d}\alpha \; , \qquad (11.32)$$

where the expression on the right-hand side represents the mean value along the circumference of the circle. Approximating the integral by a sum and considering only four points in the circle, we obtain

$$\Delta g(0, 0) = \frac{1}{4}\left[\Delta g(0, h) + \Delta g(h, 0) + \Delta g(0, -h) + \Delta g(-h, 0)\right] ;$$

see Fig. 11.6 (the z-axis is now positive downwards). This yields the formula for the downward analytic continuation in the form

$$\Delta g(0, h) = 4\Delta g(0, 0) - \left[\Delta g(0, -h) + \Delta g(-h, 0) + \Delta g(h, 0)\right] . \qquad (11.33)$$

It is evident that, at first, we must calculate value $\Delta g(0, -h)$ by analytic continuation upward.

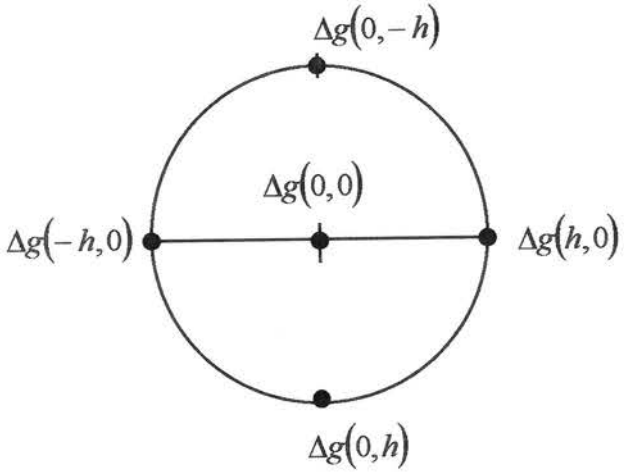

Fig. 11.6: Downward analytic continuation for a two-dimensional field.

11.3.2 First and second derivatives of the gravity acceleration

The resolving power of residual anomalies is frequently insufficient to solve geological problems, in particular to separate individual geological bodies. The horizontal gradient $\partial(\Delta g)/\partial x$, first vertical gradient $\partial(\Delta g)/\partial z$ and especially the second vertical gradient $\partial^2(\Delta g)/\partial z^2$ of gravity anomalies have better properties for this purpose. Since direct measurements of these derivatives would be complicated and uneconomical, many approximate methods have been developed for their computation by transforming anomalies Δg (Mareš et al., 1979; Pick et al., 1973).

11.3.3 Linsser's method of determining steeply inclined density discontinuities

Horizontal discontinuities in the Earth can easily be studies by seismic reflection methods. These methods, however, fail if the discontinuity is vertical, or has a large slope, because seismic waves, reflected at such a discontinuity, do not return to the surface. On the other hand, the interpretation of gravity data can contribute significantly to the investigation of such vertical discontinuities.

Steeply inclined density discontinuities or faults usually have the gravity effect of a step. To investigate such discontinuities, it is convenient to use the method of horizontal gradients, or the special method proposed by Linsser.

At an initial point, we determine the direction of the maximum horizontal gradient. Along this direction we then gradually displace a horizontal material half-plane or a vertical step of a chosen depth, but of an unknown surface density σ. In the case of the vertical step, this density represents the product $\sigma = \rho d$, where ρ is the volume density and d is the thickness (height) of the step. Density σ, as well as the deviation between the observed and theoretical

fields, may be determined by the method of least squares. The least deviation along the profile then indicates the position of the discontinuity.

In the modification by Linsser, two areas, A_1 and A_2, are first computed for each position P_i along the profile: A_1 is the area between curve $y = \Delta g(x)$ along the profile and horizontal straight line $y = \Delta g(P_i)$; A_2 is the area between the observed anomaly $\Delta g(x)$ and the theoretical vertical acceleration $\gamma(x)$ for the step (the edge of the step is situated under point P_i). Thus,

$$A_1 = \int \left| \Delta g(x) - \Delta g(P_i) \right| \mathrm{d}x \ ,$$

$$A_2 = \int \left| \Delta g(x) - \gamma(x) \right| \mathrm{d}x \ .$$

The coincidence of the observed anomalies with the theoretical values is then estimated by the parameter

$$C = \frac{A_1 - A_2}{A_1} \ .$$

Product $C\sigma$ is then computed for all points P_i. The point, where this product attains the maximum value, indicates a point of the discontinuity.

The computation is usually performed for 2 or 3 different depths of the model. Under favourable circumstances, the shifts of the solutions for different depths may indicate the slope of the density discontinuity.

Linsser's method has been applied successfully by many authors. Further details and applications of this method can be found in Mareš et al. (1979).

11.3.4 Determination of the total differential mass

Consider a perturbing body of a finite volume. Choose a point O on the Earth's surface, e.g., somewhere above the perturbing body. Construct a closed surface S, as shown in Fig. 11.7, which contains the perturbing body in its interior. Surface S consist of two parts, S_1 and S_2, where S_2 is a hemispherical surface of radius R with its centre at point O, and S_1 is the corresponding circular part of plane $z = 0$. According to Gauss' theorem,

$$\iint_S (\Delta g)_n \, \mathrm{d}S = -4\pi GM \ , \tag{11.34}$$

where $(\Delta g)_n$ is the component of the gravity anomaly in the direction of the outward normal to surface S, and M is the total differential mass. First, replace the integral in (11.34) by the sum of two integrals over S_1 and S_2:

$$\iint_{S_1} (\Delta g)_n \, \mathrm{d}S + \iint_{S_2} (\Delta g)_n \, \mathrm{d}S = -4\pi GM \ . \tag{11.35}$$

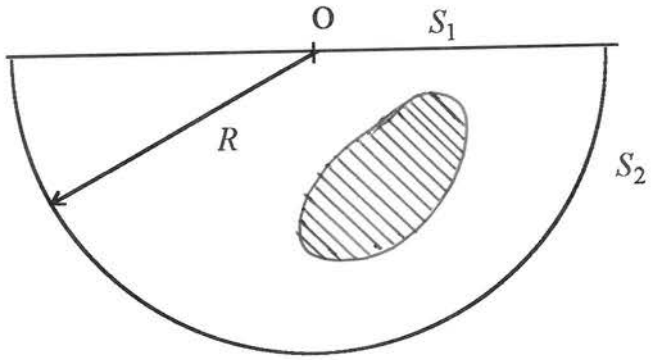

Fig. 11.7: Closed surface for the application of Gauss' theorem.

Now, let us consider the integral over hemispherical surface S_2. According to the theorem on multipole expansion, the gravitational field on this surface may be expressed as a sum of the fields of a mass point of mass M, of a dipole and higher multipoles, which are all located at point P. Thus,

$$\iint_{S_2} (\Delta g)_n \, \mathrm{d}S = \iint_{S_2} \left[-\frac{GM}{R^2} + \frac{f(\vartheta, \lambda)}{R^3} + \ldots \right] \mathrm{d}S = \tag{11.36}$$

$$= -2\pi GM + \frac{1}{R^3} \iint_{S_2} f(\vartheta, \lambda) \, \mathrm{d}S + \ldots ,$$

where f is a bounded function (it is a spherical function of the first degree), and we have taken into account that the area of surface S_2 is $2\pi R^2$. In the limiting process in which R tends to infinity, the area of surface S_2 increases as R^2, but the field of the dipole decreases as $1/R^3$, and the fields of the higher multipoles decrease even more rapidly. Consequently, these terms vanish, and only the first term remains on the right-hand side of (11.36).

The same limiting process extends the integration over S_1 on the whole plane $z = 0$. In the limit, Gauss' theorem (11.35) yields

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\Delta g)_n \, \mathrm{d}x \, \mathrm{d}y - 2\pi GM = -4\pi GM \ .$$

By changing the positive direction of the gravity acceleration downwards, as usual, i.e. by putting $\Delta g(x, y, 0) = -(\Delta g)_n$, we arrive at

$$\boxed{M = \frac{1}{2\pi G} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Delta g(x, y, 0)\, \mathrm{d}x\, \mathrm{d}y} \,. \tag{11.37}$$

This is the resultant formula for calculating the total differential mass of the perturbing body (or bodies) on the basis of the gravity anomalies on the Earth's surface. In practice, the integration is performed over a limited region only.

Note that formula (11.37) makes it possible to determine the differential mass uniquely, and this determination is independent of the shape or position of the perturbing body.

Methods of determining other general characteristics of the perturbing body, such as the coordinates of the centre of mass and others, are described in the specialised literature (Pick et al., 1973).

11.4 Interpretation

The quantitative interpretation of observed anomalies consists in a comparison of these anomalies with theoretical anomalies computed for certain models. As a first step, the corresponding forward problems must be solved.

11.4.1 Forward problem for a general three-dimensional body

Introduce a Cartesian coordinate system, the origin of which is located on the Earth's surface, the x- and y-axes are horizontal, z-axis is vertical and positive downwards (Fig. 11.8a). Calculate the gravitational field of a body of volume V at point P. Denote the coordinates of the points of the body by ξ, η, ζ, and the coordinates of the point of observation P by x, y, z. Further, denote the vector between a point Q of the body and point P by

$$\mathbf{R} = (x - \xi, y - \eta, z - \zeta) \,. \tag{11.38}$$

Its magnitude is

$$R = |\mathbf{R}| = \sqrt{(x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2} \,. \tag{11.39}$$

The intensity of the gravitational field (gravitational acceleration) at point $P(x, y, z)$ can then be expressed as

$$\mathbf{E} = -G \iiint_V \rho \frac{\mathbf{R}}{R^3} \mathrm{d}V \,, \tag{11.40}$$

where $\rho = \rho(\xi, \eta, \zeta)$ is the density, and the integration is performed with respect to coordinates ξ, η, ζ.; see analogous formulae in Chap. 7. The vertical component of intensity then reads

$$E_z(x, y, z) = -G \iiint_V \rho \frac{z - \zeta}{\left[(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2\right]^{3/2}} \mathrm{d}\xi \mathrm{d}\eta \mathrm{d}\zeta \quad .(11.41)$$

If ρ is the differential density (deviation from the normal density), this vertical component represents the theoretical gravity anomaly, $E_z = \Delta g$.

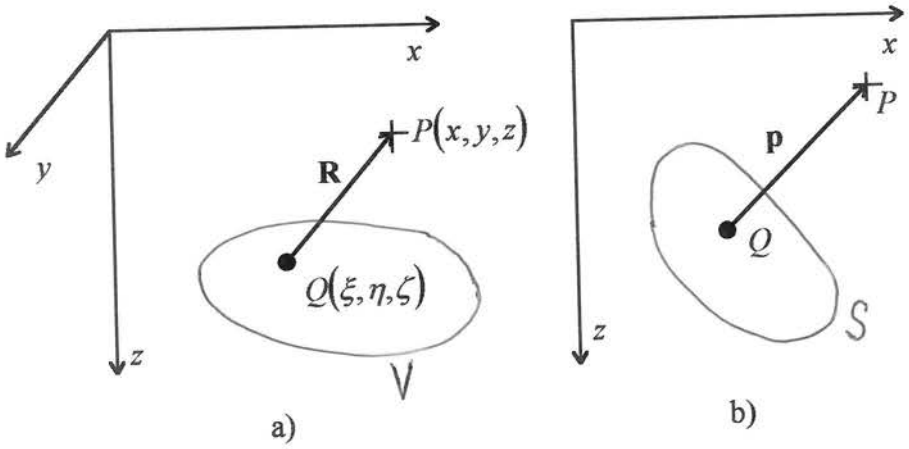

Fig. 11.8: Point of observation P and a gravitating body: a) three-dimensional body; b) cross-section through a "two-dimensional" body.

Formula (11.41) is the initial formula for computing the gravitational effect of a given body.

11.4.2 Forward problem for a "two-dimensional" body

In nature we frequently find perturbing bodies which are elongated horizontally in a single direction. The gravitational field of such a body at places near its centre will not change very much if the body is extended even farther along this direction up to infinity. In this way we arrive at the notion of a "two-dimensional" body.

A "two-dimensional" body is defined as a body which is infinitely long in one direction, say along the y-axis, and its properties are the same along each straight line which is parallel to this axis. In other words, the density in a "two-dimensional" body is independent of the y-coordinate, i.e. $\rho = \rho(x, z)$. The properties of the body are thus completely described by its properties in any cross-section which is perpendicular to the y-axis. This substantiates the use of the term "two-dimensional" body.

For a "two-dimensional" body, formula (11.41) takes the form

$$E_z(x,y,z) = -G\iint_S \int_{\eta=-\infty}^{+\infty} \rho(\xi,\zeta)\frac{(z-\zeta)\,\mathrm{d}\xi\,\mathrm{d}\zeta\,\mathrm{d}\eta}{\left[(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2\right]^{3/2}}, \qquad (11.42)$$

where S is the cross-section of the body, perpendicular to the y-axis and passing through point P. Since density ρ is now independent of variable η, the integration with respect to this variable can be performed analytically.

First, consider a simpler integral

$$I = \int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{\left(1+x^2\right)^{3/2}},$$

which can be solved by substituting $x = \tan u$. As

$$1+x^2 = 1/\cos^2 u, \quad \mathrm{d}x = \mathrm{d}u/\cos^2 u,$$

we obtain

$$I = \int_{-\pi/2}^{\pi/2} \cos u\,\mathrm{d}u = 2.$$

Therefore, express integral (11.42) as

$$E_z(P) = -G\iint_S \rho\frac{(z-\zeta)\,\mathrm{d}\xi\,\mathrm{d}\zeta}{\left[(x-\xi)^2+(z-\zeta)^2\right]^{3/2}} \int_{-\infty}^{+\infty} \frac{\mathrm{d}\eta}{\left[1+\dfrac{(y-\eta)^2}{(x-\xi)^2+(z-\zeta)^2}\right]^{3/2}}, \qquad (11.43)$$

and introduce a new variable u, instead of η, by substituting

$$\frac{y-\eta}{\sqrt{(x-\xi)^2+(z-\zeta)^2}} = \tan u. \qquad (11.44)$$

After integrating, we arrive at the general formula for the "two-dimensional" body in the form

$$E_z(P) = -2G\iint_S \rho(\xi,\zeta)\frac{z-\zeta}{p^2}\,\mathrm{d}S, \qquad (11.45)$$

where

$$p = \sqrt{(x-\xi)^2+(z-\zeta)^2} \qquad (11.46)$$

is the distance of the point of observation P from the corresponding point of the cross-section (Fig. 11.8b), and $\mathrm{d}S = \mathrm{d}\xi\,\mathrm{d}\zeta$.

An analogous formula can be written for component E_x, only the term $(z-\zeta)$ in (11.45) must be replaced by $(x-\xi)$. The intensity for a "two-dimensional" body can then be expressed in vectorial form as

$$\boxed{\mathbf{E}(P) = (E_x, 0, E_z) = -2G\iint_S \rho \frac{\mathbf{p}}{p^2}\,\mathrm{d}S}\,, \tag{11.47}$$

where vector

$$\mathbf{p} = (x-\xi, 0, z-\zeta)\ ; \tag{11.48}$$

see Fig. 11.8b.

Although we shall not use the potential for "two-dimensional" bodies, let us derive this potential, since the problem is interesting from a broader physical point of view (it is a source of many misunderstandings in some textbooks of physics).

Therefore, let us attempt to express vector $(\mathbf{p}/p^2)$ as the gradient of a scalar function. By differentiating (11.46) we get

$$\operatorname{grad} p = \left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}, \frac{\partial p}{\partial z}\right) = \left(\frac{x-\xi}{p}, 0, \frac{z-\zeta}{p}\right) = \frac{\mathbf{p}}{p}\,,$$

$$\operatorname{grad}(\ln p) = \frac{1}{p}\operatorname{grad} p = \frac{\mathbf{p}}{p^2}\ .$$

Thus,

$$\frac{\mathbf{p}}{p^2} = \operatorname{grad}(\ln p) = -\operatorname{grad}\left(\ln\frac{1}{p}\right)\ . \tag{11.49}$$

Insert this into (11.47) and interchange the order of the integration and differentiation. The intensity can then be expressed as

$$\mathbf{E} = -\operatorname{grad} V\ , \tag{11.50}$$

where

$$\boxed{V = -2G\iint_S \rho \ln\frac{1}{p}\,\mathrm{d}S} \tag{11.51}$$

is the so-called *logarithmic potential*.

We remind the reader that the corresponding gravitational potential for a "usual" body is

$$V = -G\iiint_V \frac{\rho}{R}\,\mathrm{d}V \ . \tag{11.52}$$

The potentials of this type, containing a reciprocal distance (i.e. $1/R$) in the integrand, are called *Newtonian potentials*. The reader may rightly ask why we have not derived the logarithmic potential (11.51) directly from the Newtonian potential (11.52) by extending the corresponding limits to $\pm\infty$. However, the problem is that the Newtonian potential (11.52) becomes infinite in this case. To prove this, it is sufficient to analyse the following simpler integral (since R is given by (11.39)):

$$I = \int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{\sqrt{1+x^2}} \ . \tag{11.53}$$

This integral can be calculated using the substitution $x = \sinh u$. Considering that

$$\cosh^2 x - \sinh^2 x = 1 \ , \quad \mathrm{d}x = \cosh u\,\mathrm{d}u \ , \tag{11.54}$$

we obtain

$$I = \int_{-\infty}^{+\infty} \mathrm{d}u = +\infty \ . \tag{11.55}$$

We have thus arrived at the conclusion that the applicability of Newtonian potentials is somewhat limited, that they can be used, roughly speaking, only for bounded bodies. This is in contrast with the general validity of formula (11.40) for intensity, which holds true for bounded, as well as unbounded bodies. It also means that the intensity is a fundamental physical quantity, whereas the potential is rather an auxiliary and derived quantity, although very convenient.

Let us remind the reader also of some electrostatic problems, where the Newtonian potential cannot be introduced, e.g., for the electrostatic field of a charged straight line, or a charged plane.

11.4.3 Forward and inverse problems for a sphere

In order to demonstrate the simplest classical methods of gravimetric interpretations, consider a homogeneous sphere. Its gravitational effect at external points is the same as if the mass of the sphere, M, were concentrated at its centre T. Let the z-axis pass through the centre of the sphere (Fig. 11.9). Denote the depth of its centre by h.

Considering the symmetry of the gravitational field of the sphere, let us study the field on the Earth's surface along the x-axis only. At point $P(x, 0, 0)$, the gravitational acceleration due to the sphere is GM/r^2, where $r = \sqrt{x^2 + h^2}$ is

the distance from the centre of the sphere. The vertical component of the acceleration is

$$E_z(x) = \frac{GM}{r^2}\cos\alpha = \frac{GMh}{r^3} = \frac{GMh}{\left(x^2+h^2\right)^{3/2}}\ , \tag{11.56}$$

α being the angle between the vertical and the direction towards the centre of the sphere. Formula (11.56) follows also immediately from (11.41) by putting $\xi = \eta = 0,\ \zeta = h,\ z = y = 0$. This vertical component represents the gravity anomaly, so that we shall also put $\Delta g(x) = E_z(x)$.

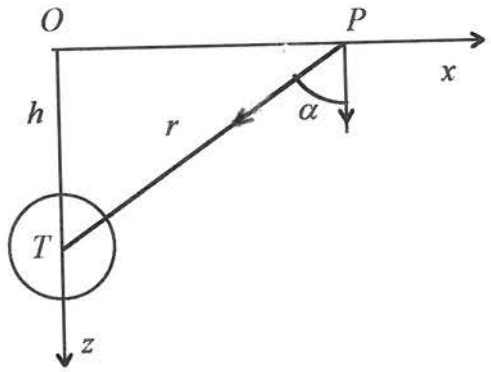

Fig. 11.9: Position of the sphere and of the point of observation, P.

The gravity anomaly (on the Earth's surface) attains its maximum value at origin O, i.e. for $x = 0$:

$$\left(\Delta g\right)_{\max} = \frac{GM}{h^2}\ . \tag{11.57}$$

This is one equation for the determination of two unknowns, M and h.

Consider another point on the x-axis, e.g., the point at which the anomaly is equal to half of its maximum value. Denote its coordinate by $x_{1/2}$. Then,

$$\frac{1}{2}\frac{GM}{h^2} = \frac{GMh}{\left(x_{1/2}^2+h^2\right)^{3/2}}\ . \tag{11.58}$$

From this equation we can determine the depth of the centre of the sphere:

$$h = \frac{\left|x_{1/2}\right|}{\sqrt{2^{2/3}-1}} = 1.305\left|x_{1/2}\right|\ . \tag{11.59}$$

The mass of the sphere can then be determined from (11.57).

We can summarise the interpretation process as follows. Assume that an observed anomaly is due to a sphere (e.g., on the basis of the circular shape of the isolines). The horizontal coordinates of the sphere can easily be determined as the coordinates of the point at which the anomaly attains its maximum value. The depth h and mass M of the sphere can then be determined from the measured values of $(\Delta g)_{\max}$ and $x_{1/2}$ by means of (11.59) and (11.57). This represents the complete solution of the inverse problem for a homogeneous sphere. The radius of the sphere cannot be determined from these measurements; the density would have to be known.

Instead of distance $x_{1/2}$, the distance of another characteristic point may be used. For example, we could use the distance of the inflectional point, where $\partial^2 E_z / \partial x^2 = 0$, or of another point. However, at present we should rather use the values of anomalies at a greater number of points and determine the parameters of the sphere by the least-squares method, using a computer.

Analogous simple methods of solving inverse gravimetric problems have been developed also for other homogeneous bodies, such as a cylinder, vertical step or belt. For details we refer the reader, e.g., to Pick et al. (1973).

11.4.4 A "two-dimensional" cylinder or a material straight line

Consider a homogeneous rotational cylinder which is infinitely long and its axis is horizontal, parallel to the y-axis at depth h; see Fig. 11.10.

The gravitational effect of the cylinder at external points is the same as if the mass of the cylinder were uniformly concentrated along its axis. The linear density of the equivalent material straight line is $\lambda = \rho S$, where ρ is the volume density in the cylinder, and S is the area of the cross-section of the cylinder.

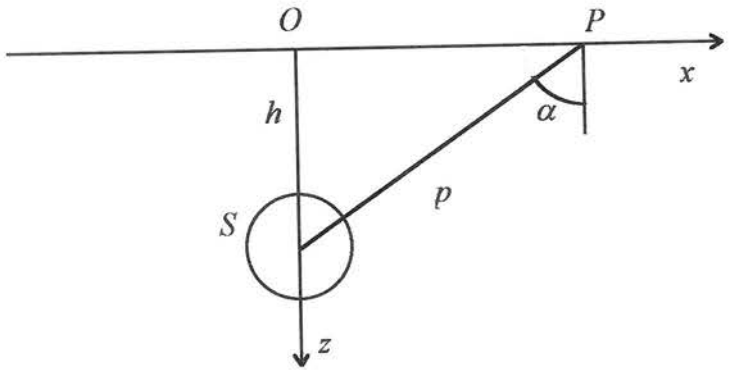

Fig. 11.10: Cross-section through an infinitely long cylinder.

It follows from (11.45) that the corresponding gravity effect along the x-axis reads

$$\Delta g(x) = E_z(x) = \frac{2G\lambda h}{x^2 + h^2} \ . \tag{11.60}$$

Let us show that this formula follows also immediately from Gauss' law

$$\iint_S E_n \, \mathrm{d}S = -4\pi GM \ .$$

Construct an auxiliary cylinder in the following way: its axis coincides with the axis of the material cylinder; its length along the y-axis, l, is arbitrary but finite; its radius is p, so that its lateral surface passes through the point of observation P. Apply Gauss' theorem to the surface of this cylinder. As the intensity is perpendicular to the axis, the surface integrals over the bases of the cylinder are zero. Consequently, only the integral over the lateral surface must be considered:

$$E_n 2\pi pl = -4\pi G\lambda l \ .$$

Thus, the radial component is $E_n = -2G\lambda/p$, and the vertical component is $E_z = -E_n \cos\alpha = -E_n h/p$, which yields (11.60).

The anomaly (11.60) is maximum for $x = 0$:

$$(\Delta g)_{\max} = \frac{2G\lambda}{h} \ . \tag{11.61}$$

Denote by $x_{1/2}$ again the distance at which the anomaly is equal to half of its maximum value. Thus,

$$\frac{1}{2}(\Delta g)_{\max} = \frac{2G\lambda h}{x_{1/2}^2 + h^2} \ .$$

Considering (11.61), we arrive at a very simple formula for the determination of the depth of the axis of the cylinder, namely

$$h = \left| x_{1/2} \right| \ ; \tag{11.62}$$

compare this formula with (11.59) for a sphere. The linear density λ can then be determined from $(\Delta g)_{\max}$ using (11.61).

11.4.5 A "two-dimensional" prism

Any "two-dimensional" body of an irregular cross-section may be approximated by a system of "two-dimensional" prisms if the irregular body is subdivided by

two systems of planes which are perpendicular to the x- and z-axes. Therefore, consider a homogeneous "two-dimensional" prism of a rectangular cross-section, the edges of which are parallel to the y-axis. Consequently, the edges of the cross-section are parallel to the x- and z-axes (Fig. 11.11). Denote their x-coordinates by ξ_1 and ξ_2, and their z-coordinates by ζ_1 and ζ_2.

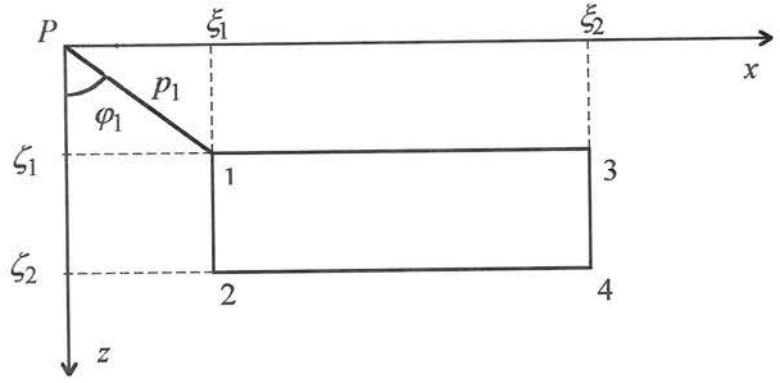

Fig. 11.11: "Two-dimensional" prism.

Without loss of generality, calculate the gravitational field of the prism at the origin of the coordinate system only, i.e. for $x = z = 0$. Formula (11.45) then yields

$$E_z(0,0) = 2G\rho \int_{\xi_1}^{\xi_2}\int_{\zeta_1}^{\zeta_2} \frac{\zeta}{\xi^2+\zeta^2}\,\mathrm{d}\xi\,\mathrm{d}\zeta \ . \tag{11.63}$$

The integration with respect to ζ is easy to perform, since

$$2\int \frac{\zeta}{\xi^2+\zeta^2}\,\mathrm{d}\zeta = \ln\left(\xi^2+\zeta^2\right) . \tag{11.64}$$

Then, by integrating by parts we obtain

$$\int \ln\left(\xi^2+\zeta^2\right)\mathrm{d}\xi = \xi\ln\left(\xi^2+\zeta^2\right) - \int \frac{2\xi^2}{\xi^2+\zeta^2}\,\mathrm{d}\xi = \tag{11.65}$$

$$= \xi\ln\left(\xi^2+\zeta^2\right) - 2\int \frac{\left(\xi^2+\zeta^2\right)-\zeta^2}{\xi^2+\zeta^2}\,\mathrm{d}\xi = \xi\ln\left(\xi^2+\zeta^2\right) - 2\xi + 2\zeta\arctan\frac{\xi}{\zeta} \ .$$

Term 2ξ will vanish after introducing the limits for ζ. Therefore,

$$E_z(0,0) = G\rho\left[\left[\xi\ln\left(\xi^2+\zeta^2\right)+2\zeta\arctan\frac{\xi}{\zeta}\right]_{\xi_1}^{\xi_2}\right]_{\zeta_1}^{\zeta_2}. \qquad (11.66)$$

For the sake of brevity, introduce the following notations:

$$p = \sqrt{\xi^2+\zeta^2}\ , \quad \varphi = \arctan\left(\xi/\zeta\right), \qquad (11.67)$$

where p is the distance from point P, and φ is the angle between the radius-vector $\mathbf{p}$ and the z-axis. Using the notation of the vertices as shown in Fig. 11.10, formula (11.66) can be expressed in the final form as follows:

$$E_z(0,0) = 2G\rho\left[\xi_2\ln\frac{r_4}{r_3} - \xi_1\ln\frac{r_2}{r_1} + \zeta_2\left(\varphi_4-\varphi_2\right) - \zeta_1\left(\varphi_3-\varphi_1\right)\right]. \qquad (11.68)$$

The solution of the inverse problem for the prism is difficult. To facilitate it, auxiliary tables were compiled (see Pick et al., 1973). It can be seen from (11.66) that this inverse problem leads to a system of transcendental equations. Such systems of equations must be solved by numerical methods.

Note that in the previous subsection we derived the gravitational field of an rotational cylinder both from the general formula (11.45) for a "two-dimensional" body, and also from Gauss' law. Here, for the gravitational field of a prism, we have used only the general formula (11.45). In this case it is not evident how to apply Gauss' law because the field does not have the necessary symmetry.

11.4.6 A rectangular parallelepiped

Since a system of rectangular parallelepipeds may be used to approximate an arbitrary three-dimensional body, the gravitational field of such an elementary body is of special interest. Therefore, consider a rectangular parallelepiped, the edges of which are parallel to the coordinate axes x, y, z (Fig. 11.12). Let its volume be delimited by planes $x = \xi_i$, $y = \eta_i$ and $z = \zeta_i$, where $i = 1, 2$.

Calculate the gravitational effect of the parallelepiped again at the coordinate origin only, i.e. at point $P = (0, 0, 0)$. Formula (11.41) then yields

$$E_z(P) = G\rho\int_{\xi_1}^{\xi_2}\int_{\eta_1}^{\eta_2}\int_{\zeta_1}^{\zeta_2}\frac{\zeta}{\left(\xi^2+\eta^2+\zeta^2\right)^{3/2}}\,\mathrm{d}\xi\,\mathrm{d}\eta\,\mathrm{d}\zeta\ . \qquad (11.69)$$

The integration with respect to the vertical coordinate ζ can easily be performed:

$$E_z(P) = G\rho \left[\int_{\xi_1}^{\xi_2} \int_{\eta_1}^{\eta_2} \frac{\mathrm{d}\,\xi\,\mathrm{d}\,\eta}{\sqrt{\xi^2+\eta^2+\zeta^2}} \right]_{\zeta_1}^{\zeta_2} . \tag{11.70}$$

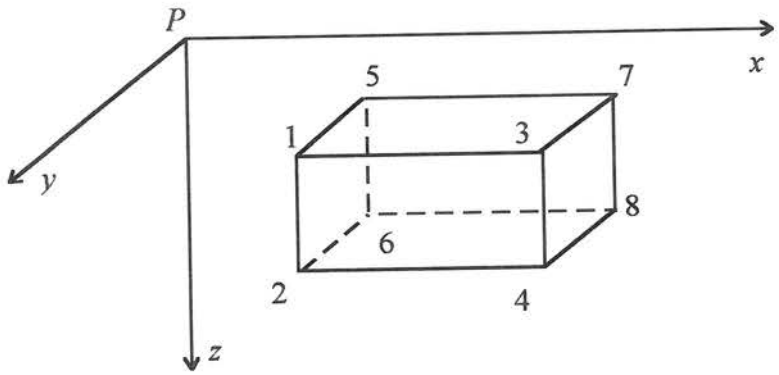

Fig. 11.12: Rectangular parallelepiped.

The integral with respect to η belongs to a category of integrals which are not usually solved in standard courses of mathematical analysis. Therefore, let us describe this integration in greater detail. Denoting

$$x = \eta \Big/ \sqrt{\xi^2+\zeta^2} \ , \tag{11.71}$$

the integration leads to the indefinite integral

$$I_1 = \int \frac{\mathrm{d}\,x}{\sqrt{1+x^2}} \ , \tag{11.72}$$

which can be solved by substituting $x = \sinh u$; see analogous formulae (11.53) and (11.54). We obtain

$$I_1 = u + C = \operatorname{arg\,sinh} x + C \ ,$$

where C is an arbitrary constant, but we shall omit the constants of integration everywhere.

Since the hyperbolic sinus is defined by the exponential function,

$$x = \sinh u = \frac{e^u - e^{-u}}{2} \ ,$$

it should be possible to express the inverse function as a logarithm. Indeed, by multiplying the previous relation by e^u, one gets a quadratic equation

$$\left(e^u\right)^2 - 2xe^u - 1 = 0$$

in the unknown e^u. Its positive solution is

$$e^u = x + \sqrt{1+x^2} \ ,$$

so that for $u = \operatorname{arg\,sinh} x$ we obtain

$$\operatorname{arg\,sinh} x = \ln\left(x + \sqrt{1+x^2}\right) . \tag{11.73}$$

Integral (11.72) can then be expressed in the following usual form:

$$\int \frac{\mathrm{d}\,x}{\sqrt{1+x^2}} = \ln\left(x + \sqrt{1+x^2}\right) . \tag{11.74}$$

By inserting (11.71) into (11.74) and adding the integration constant $\ln\sqrt{\xi^2+\zeta^2}$, one gets another useful formula

$$\int \frac{\mathrm{d}\,\eta}{\sqrt{\xi^2+\eta^2+\zeta^2}} = \ln\left(\eta + \sqrt{\xi^2+\eta^2+\zeta^2}\right) . \tag{11.75}$$

The reader may verify formulae (11.74) and (11.75) by differentiation. Formula (11.70) now takes the form

$$E_z(P) = G\rho\left[\left[\int_{\xi_1}^{\xi_2} \ln\left(\eta + \sqrt{\xi^2+\eta^2+\zeta^2}\right)\mathrm{d}\,\xi\right]_{\eta_1}^{\eta_2}\right]_{\zeta_1}^{\zeta_2} . \tag{11.76}$$

Finally, consider the indefinite integral

$$I_2 = \int \ln\left(\eta + \sqrt{\xi^2+\eta^2+\zeta^2}\right)\mathrm{d}\,\xi \ . \tag{11.77}$$

After integrating by parts, we get

$$I_2 = \xi\ln(\eta + r) - \int \frac{\xi^2\,\mathrm{d}\,\xi}{(\eta+r)r} \ , \tag{11.78}$$

where we have denoted

$$r = \sqrt{\xi^2 + \eta^2 + \zeta^2} \ . \tag{11.79}$$

The second term on the right-hand side of (11.78) can be rearranged by substituting

$$t = \xi + r = \xi + \sqrt{\xi^2 + \eta^2 + \zeta^2} \ , \tag{11.80}$$

and the modified integral can be solved in the usual way (Pick, 1984). However, the integral obtained is rather complicated. Therefore, we shall first rearrange (11.78).

It can be seen that variables ξ and η in the integrand of (11.70) appear quite symmetrically. If the integration with respect to ξ were performed first, integral (11.78) would contain the symmetrical term, i.e. $\eta \ln(\xi + r)$. Thus, add and subtract this term on the right-hand side of (11.78), but express the subtracted term as the integral of the corresponding derivative, i.e. as $\int (\eta / r) \mathrm{d}\xi$. Integral (11.78) now becomes

$$I_2 = \xi \ln(\eta + r) + \eta \ln(\xi + r) - I_3 \ , \tag{11.81}$$

where

$$I_3 = \int \left[\frac{\xi^2}{(\eta + r) r} + \frac{\eta}{r} \right] \mathrm{d}\xi = \int \frac{\xi^2 + \eta^2 + \eta r}{r(\eta + r)} \mathrm{d}\xi \ . \tag{11.82}$$

The integrand in (11.82) may be simplified by adding and subtracting ζ^2 in the numerator. Considering (11.79) we have

$$I_3 = \int \left[1 - \frac{\zeta^2}{(\eta + r) r} \right] \mathrm{d}\xi = \xi - \zeta^2 \int \frac{\mathrm{d}\xi}{r(\eta + r)} \ .$$

The last integral can be solved by substituting (11.80):

$$\begin{aligned} I_3 &= \xi - \zeta^2 \int \frac{2\,\mathrm{d}t}{\eta^2 + \zeta^2 + 2\eta t + t^2} = \xi - 2 \int \frac{\mathrm{d}t}{1 + \left(\dfrac{\eta + t}{\zeta} \right)^2} = \\ &= \xi - 2\zeta \arctan \frac{t + \eta}{\zeta} = \xi - 2\zeta \arctan \frac{\xi + \eta + r}{\zeta} \ . \end{aligned} \tag{11.83}$$

By using (11.83) and (11.81), integral (11.77) can be expressed in the final form as

$$\int \ln(\eta+r)\,\mathrm{d}\xi = \xi\ln(\eta+r)+\eta\ln(\xi+r)-\xi+2\zeta\arctan\frac{\xi+\eta+r}{\zeta}\,, \quad (11.84)$$

where r is given by (11.79).

The term $(-\xi)$ in (11.84) does not change if the upper and lower bounds are substituted for ζ and η. Consequently, this term will vanish in the definite integral. By taking this fact into account and inserting (11.84) into (11.76), we arrive at the resultant formula for the gravity effect of the homogeneous rectangular parallelepiped at the origin of the coordinate system:

$$E_z(P) = G\rho\left[\left[\left[\xi\ln\left(\eta+\sqrt{\xi^2+\eta^2+\zeta^2}\right)+\eta\ln\left(\xi+\sqrt{\xi^2+\eta^2+\zeta^2}\right)+\right.\right.\right. \quad (11.85)$$

$$\left.\left.\left.+2\zeta\arctan\frac{\xi+\eta+\sqrt{\xi^2+\eta^2+\zeta^2}}{\zeta}\right]_{\xi_1}^{\xi_2}\right]_{\eta_1}^{\eta_2}\right]_{\zeta_1}^{\zeta_2}.$$

This is a special case of a more general formula which was derived by Pick (1984). For computational purposes, however, formula (11.85) must be rearranged; in particular, it is necessary to insert the corresponding bounds there and to simplify some expressions. For example, the difference

$$\xi_2\left[\ln\left(\eta_2+\sqrt{\xi_2^2+\eta_2^2+\zeta_2^2}\right)-\ln\left(\eta_2+\sqrt{\xi_2^2+\eta_2^2+\zeta_1^2}\right)\right] \quad (11.86)$$

should be expressed as

$$\xi_2\ln\frac{\eta_2+\sqrt{\xi_2^2+\eta_2^2+\zeta_2^2}}{\eta_2+\sqrt{\xi_2^2+\eta_2^2+\zeta_1^2}}\,, \quad (11.87)$$

etc. The latter formula is not only faster, since only one logarithmic function must be calculated, but also more accurate. Namely, expression (11.86) for a distant parallelepiped is computed as the difference between two close large numbers. This difference is accompanied by a loss of significant figures. Also from this point of view, expression (11.87) is more convenient.

11.4.7 Some comments on the interpretation of gravity anomalies

There are now many solutions to the forward and inverse gravimetric problems so that only the fundamental features of some of them need be mentioned.

Numerous formulae for the gravity effect of other homogeneous bodies can be found, e.g., in Mareš et al. (1979), Pick et al. (1973), Talwani (1973). The gravity effect of bodies with a variable density was studied by Pick (1984).

As regards the three-dimensional bodies, the forward and partly also the inverse gravimetric problems have been solved for ellipsoids, paraboloids and finite cylinders, among others.

As regards "two-dimensional" bodies, especially the method by Talwani et al. (1959) has frequently been used. This method yields an analytical formula for the gravity effect of homogeneous bodies which are bounded by a general polygon:

$$E_z(0,0) = 2G\rho \sum_{i=1}^{N} \frac{x_i z_{i+1} - z_i x_{i+1}}{(x_{i+1} - x_i)^2 + (z_{i+1} - z_i)^2} \cdot \left[(x_{i+1} - x_i)\left(\arctan \frac{x_{i+1}}{z_{i+1}} - \arctan \frac{x_i}{z_i} \right) + \frac{1}{2}(z_{i+1} - z_i) \ln \frac{x_{i+1}^2 + z_{i+1}^2}{x_i^2 + z_i^2} \right], \tag{11.88}$$

where ρ is the differential density and (x_i, z_i) are the coordinates of the vertices of the polygon (Švancara, 1986).

Ever increasing attention has recently been paid to the so-called "two and a half dimensional" models ($2\frac{1}{2}$–D models), which are "two-dimensional" models, but of a finite length and bounded perpendicularly to the longitudinal axis of the body. The conception of $2\frac{1}{2}$–D bodies makes it possible to preserve the simplicity of two-dimensional modelling and, to a large extent, to describe the effects of many there-dimensional bodies. Suitable formulae were derived, e.g., by Rasmussen and Pedersen (1979), and their numerical properties analysed by Švancara (1986).

The inverse gravimetric problems for complicated structural models usually lead to large systems of linear or non-linear equations which are usually ill-posed (i.e. the corresponding determinant is close to zero or equal to zero) or even underdetermined (the number of equations is less than the number of unknown parameters of the model). The solutions of such systems are unstable or even not uniquely defined. To reduce the ambiguity of the solution, all available a priori data should be taken into account, such as measurements in boreholes, seismic data, geological information, etc. Moreover, stable algorithms for solving systems of equations should be preferred, such as various descent methods (the method of steepest descent, the method of conjugate gradients, or some others). Detailed descriptions of the methods of solving systems of linear or non-linear equations can be found in textbooks on numerical methods or inverse problems. We also intend to include a special chapter on inverse problems in the lecture notes on seismic surface waves (in preparation).

Chapter 12

Satellite Methods of Studying the Gravitational Field - An Elementary Theory

Satellite methods have extended tremendously our knowledge of the gravitational field and figure of the Earth. Satellites may be employed in a number of ways in gravimetry and geodesy. Let us mention some of them:

- A satellite can be used as a passive object in *geometrical methods* of space geodesy; i.e. as a vertex of a triangle in space triangulation. The angles can be determined by synchronously photographing the satellite on the background of stars from different places on the Earth's surface. Angular measurements are now usually being replaced by laser measurements of distances. We shall not deal with these methods here.
- By measuring the trajectory of a satellite, we can study the forces acting on the satellite. Optical measurements, the radio Doppler effect or laser telemetry have been used for these purposes. The principles and basic theory of these *dynamical methods* will be described in this chapter and in the chapter which follows.
- *Remote sensing* from a satellite whose orbit is known with a high accuracy. Among these methods, satellite altimetry should be emphasised in particular.

We have divided the treatment of satellite methods into two separate chapters. In this chapter we solve some simpler problems using Newtonian mechanics only, whereas the treatment in the next chapter (Chapter 13) is based on methods of analytical mechanics.

In the first three sections of this chapter we shall briefly compare the main methods of investigating the gravitational field outside the Earth. Special attention will be paid to satellite methods and to the description of satellite orbits.

The remaining sections of this chapter and the next chapter will be devoted to calculations of planetary and satellite orbits. We shall consider more and more complicated problems, and try to solve them by applying various physical methods with gradually increasing complexity. In particular:

a) *Circular orbits* are calculated by means of Newtonian mechanics (Sections 12.4 and 12.5). We consider not only the elementary case of a circular orbit in a central gravitational field, but also perturbations of the circular orbit due to the Earth's flattening (the formula for the regression of the orbital plane is derived). We think that the latter problem could even serve as an extension of classical courses of physics at faculties of physics.

b) *Elliptical orbits* in a central field, i.e. the so-called Kepler problem, are studies using elementary methods in this chapter, and using Lagrange's

equations in Chapter 13. The formulation of this problem in polar coordinates can be found in many textbooks, but we shall also describe the complete derivation of Kepler's laws.

c) *Perturbations* of elliptical orbits due to deviations of the gravitational field from spherical symmetry are studied by means of the Hamilton-Jacobi theory in Chapter 13. As a special case, the effect of the Earth's flattening is again considered.

In the last section of Chapter 13, some results of satellite investigations of the Earth's gravitational field are briefly summarised.

The reader who is not interested in theoretical problems of satellite motions may skip the corresponding parts, e.g., Section 12.5 and some others.

12.1 Methods of Studying the Earth's Gravitational Field

In this chapter we shall return to the description of force fields in an inertial reference frame. We shall not consider the centrifugal force due to the Earth's rotation about its axis and, therefore, we shall speak only of the gravitational field, not of the gravity field. Of course, we shall again restrict ourselves to investigations of the external gravitational field (the field outside the Earth). For studies of the internal field it would be necessary to know the density distribution within the Earth.

There are two principal methods of studying the external gravitational field:

1. *Solutions of boundary-value problems* of the potential theory. Since the external gravitational potential satisfies Laplace's equation, it is sufficient, in principle, to know the figure of a closed external surface and certain boundary values on it.
2. Investigations of the external field by means of *satellite methods*.

Let us characterise briefly the advantages and disadvantages of these methods.

In solving the boundary-value problems for the external gravitational field of the Earth, we encounter serious problems and limitations. For example, the geoid does not satisfy the conditions to be applied as the corresponding boundary surface, since some masses exist above the geoid, its shape is not known accurately (see Chapter 14), and the gravity measurements do not cover the whole of the Earth's surface.

As opposed to this, satellite methods have some advantages. Observations of artificial satellites have yielded the most accurate and most complete data on the main features of the Earth gravitational field. Moreover, satellite altimetry has provided us with the most detailed data on the shape of the geoid over the seas. However, on the continents, satellite measurements and ground gravity measurements must be combined to complement one another. Satellite data contain information on features of larger scales, whereas ground data provide information on details of the gravity field.

The Earth's gravitational field can be studied directly by measuring the accelerations exerted on satellites. At present, such measurements are possible

and are being carried out. Nevertheless, we try to avoid such direct methods, if possible, for several reasons:

a) The acceleration is predominantly due to the first term of the potential series, i.e. to the term $(-GM/r)$. However, we are also interested in the other terms of the potential series, which do not exceed one thousandth of the main term $(-GM/r)$. In other words, we also want to study the details of the gravitational field, which are hidden in the field of the first term. For this reason, we shall usually prefer other methods (measurements of perturbations) which make it possible to separate the effect of the first term.
b) Rather a large and irregular deceleration is due to atmospheric drag. This effect must be excluded from the measured accelerations. Luckily, there are orbital parameters which are less affected by atmospheric drag. Studies of such parameters should be preferred for our purposes.
c) Direct measurements of the satellite acceleration require a network of observatories with accurately known positions. Such networks are now available, but they did not exist at the beginning of the satellite era.

As a consequence of these problems, the first satellite investigations were based predominantly on classical astronomical methods, which study the elements of orbits and their perturbations. We shall deal mainly with these classical methods.

12.2 Principles of Satellite Methods. Kepler's Laws

If the distribution of masses within the Earth were exactly *spherically symmetrical* and if there were no other influences, such as atmospheric drag or the effects of other celestial bodies, the motion of an artificial satellite would satisfy *Kepler's laws* exactly (as applied to the motion of the satellite).

Assume that the effects of atmospheric drag and of other celestial bodies have been calculated and introduced in the form of corrections. By determining the *deviations in the satellite motion from Kepler's laws*, we can then determine *deviations of the gravitational potential from spherical symmetry*, i.e. determine other terms in the potential series.

The three *Kepler laws*, formulated for the motion of the planets, state:

I. All planets move in elliptical paths, with the Sun at one focus.
II. A line from the Sun to a planet sweeps through equal areas in equal time.
III. The squares of the periods of revolution of the planets about the Sun are proportional to the cubes of the semi-major axes of the ellipses,

$$\frac{T_1^2}{T_2^2} = \frac{a_1^3}{a_2^3} \,. \tag{12.1}$$

Later we shall consider only the effects of the gravitational field of the Earth (or of another central body) on the satellite orbit. Hence, a brief discussion of the other effects should be given here.

The drag due to the Earth's atmosphere causes that the orbit becomes gradually more circular and approaches the Earth (its eccentricity and semi-major axis decrease with time). If the height of a satellite above the Earth's surface has decreased below about 200 km, atmospheric drag becomes so strong that the satellite, after several revolutions, will burn up in the lower layers of the atmosphere.

The gravitational effects of the other celestial bodies can be calculated, since the masses and positions of these bodies are well known.

Among the non-gravitational effects, also the pressure of the Sun's radiation must be taken into account. The radiation pressure played an important role mainly in the case of balloon satellites, which were used for geodetic purposes. To reduce this effect, the present geodetic satellites are constructed as massive and much smaller bodies.

The velocities of satellites, reaching nearly $8\ \mathrm{km\,s^{-1}}$, are much higher than any other macroscopic velocities in a laboratory. The question, therefore, arises whether relativistic effects are important. We can say that relativistic effects are still small in comparison with other effects. Therefore, the common practice in celestial mechanics is such that trajectories are calculated using Newtonian mechanics (supplemented by methods of analytical mechanics), and only small relativistic corrections are added, if necessary. This approach is much simpler than to perform all the calculations in the framework of relativistic mechanics.

12.3 Orbital Elements

The motion of a mass point is described by Newton's second law. This law represents three scalar equations of the second order. Their general solution thus contains six integration constants. The initial coordinates, x_0, y_0, z_0, and initial velocities, $\dot{x}_0, \dot{y}_0, \dot{z}_0$, are usually adopted as these integration constants. The motion of a mass point is described completely if the forces acting on it, its initial position and initial velocity are known.

This specification of trajectories is not suitable in case of periodic motions, which are encountered in celestial mechanics. Instead of the initial position and initial velocity, another set of parameters, the so-called orbital elements, is used in classical celestial mechanics.

In the previous chapters we described the gravitational field of the Earth in a non-inertial reference system which was fixed with the rotating Earth. This was quite natural, since the observer on the Earth's surface and his gravimeter rotate together with the Earth. We, therefore, spoke of the gravity field. On the other hand, the satellite is not immediately bound to the Earth. Its orbit is nearly constant in an inertial system, and the Earth rotates below the orbit. Therefore, it will be more appropriate to study the motion of the satellite in an inertial

system. Consequently, in this chapter and in the next chapter we shall deal only with the gravitational field of the Earth, and we shall not speak of the gravity field.

12.3.1 Elements of a satellite orbit

The plane of the Earth's equator is generally used as the basic reference plane in determining the orbital elements of artificial satellites of the Earth. The origin of the coordinate system is taken to be at the mass centre of the Earth. The vernal equinox is usually used as the basic reference point, although any other point in the equatorial plane, fixed in space, could also be used. We shall regard this reference system as an approximation to an inertial system. Nevertheless, several comments should be added.

The reference system, just defined, might be regarded as an exact inertial system if there were no other bodies in the Solar System, especially the Moon and the Sun. The presence of these bodies causes the Earth's centre to move, and the equatorial plane to change its orientation slowly as a result of precession and nutation. Therefore, our coordinate system should actually be regarded as a slightly non-inertial system. Let us discuss its motions separately:

1. The Earth and another celestial body (the Sun and the Moon) revolve round the common centre of mass. In such a rotating system, the gravitational force of the body and the corresponding centrifugal force partly compensate each other, giving rise to a residual force, i.e. the tidal force (see Chapter 5). Consequently, *tidal forces* should be added into the equations of motion of the satellite.
2. Since the satellite is not static, but moves in this system, the corresponding *Coriolis force* should be considered as well.
3. The effect of *precession and nutation* could be removed by introducing small corrections for these motions (both for the motion of the reference plane, and for the additional gravitational perturbation of the satellite orbit due to the precessing central body).

To a certain extent we could avoid these problems if, for example, the ecliptical plane were adopted as the reference plane. However, this would lead to other complications, since the series for the gravitational potential is expressed in a reference system which is fixed with the Earth.

Hereafter in this chapter we shall neglect the gravitational effects of the other celestial bodies and assume the coordinate system to be inertial. We shall also assume for a moment that the satellite moves precisely in accordance with Kepler's laws, i.e. its orbit is an ellipse which is fixed in an inertial system.

Consider two planes as shown in Fig. 12.1, namely the plane of the Earth's equator, and the plane of the satellite's orbit (actually, the orbit itself is shown there). The two planes intersect along a straight line, called the *nodal line*. The following points are marked in the figure: E is the Earth's centre, S is the position of the satellite, P is the perigee, i.e. the point of the orbit closest to the Earth, A is the apogee, i.e. the most distant point of the orbit, X the vernal

equinox (or another fixed point in space). The point where the satellite crosses the equatorial plane, on moving from the Southern to the Northern Hemisphere, is called the *ascending node*, N_a. The situation in the orbital plane is shown in Fig. 12.2.

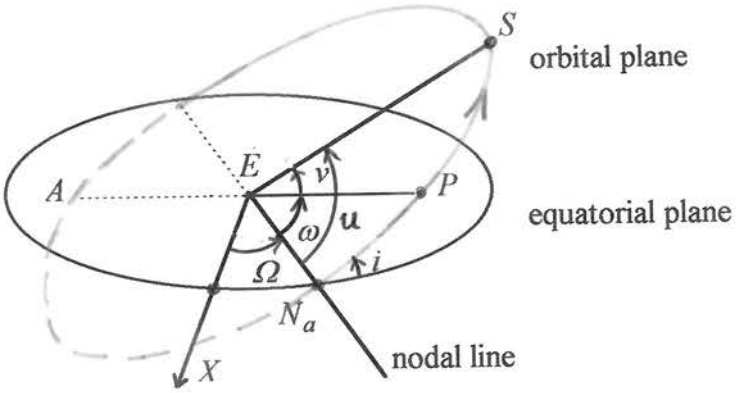

Figure 12.1: A satellite orbit in space.

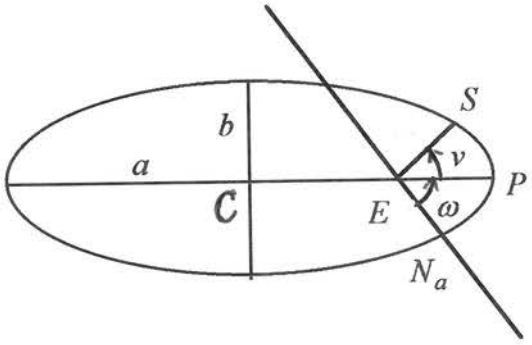

Figure 12.2: Situation in the orbital plane.

The *orbital elements*, used in classical celestial mechanics, are as follows:

Ω – the *ascension (longitude) of the ascending node*. This is the angle between the direction from the Earth's centre towards the vernal equinox and the nodal line.

i – the *inclination* of the orbital plane to the equatorial plane.

ω – the *argument of the perigee*, which is the angular distance of perigee P from ascending node N_a.

a – the *semi-major axis* of the orbit.

e – the *eccentricity* of the orbit (see the explanation below).

v – the *true anomaly* (true latitude), which is the angular distance of the satellite S from perigee P. This may be replaced by angle $u = \omega + v$, which is called the *argument of the latitude*. This is the angular distance of the satellite from ascending node N_a. As opposed to the previous orbital elements, angles v and u are functions of time. Consequently, certain constants are also introduced instead of them (see below).

Note that in geometry, the eccentricity of an ellipse is usually defined as the distance of the focus from the centre of the ellipse, $e^* = \overline{CE} = \sqrt{a^2 - b^2}$, where a, b are the semi-major and semi-minor axes, respectively. The numerical eccentricity is then introduced as the ratio $e = e^*/a$. In astronomy this dimensionless quantity e is understood to be the eccentricity.

Angles Ω and i determine the position of the orbital plane in space. Angle ω describes the orientation of the orbit in this plane, and parameters a and e describe the size and shape of the orbital ellipse. Angles v or u determine the instantaneous position of the satellite in this ellipse.

The six orbital elements, Ω to v, determine the position of the satellite uniquely, but they do not have the character of the six initial conditions yet, because angle v is a function of time. Therefore, we should introduce another constant instead of v. One possibility is to use the time of the passage through the perigee, t_0, i.e. the time at which $v = 0$. The six orbital elements Ω, i, ω, a, e, t_0, together with the given time t, then also define the satellite's position uniquely. Another possibility is to use the instantaneous mean anomaly v_M. It replaces quantities v or u, which do not change uniformly with time as a consequence of Kepler's second law (they are linear functions of time only for circular orbits). Introduce the mean motion (mean angular orbital velocity),

$$n = \frac{2\pi}{T}\,, \tag{12.2}$$

where T is the orbital period. The mean anomaly is then defined as

$$v_M = v_0 + n\left(t - \bar{t}_0\right)\,, \tag{12.3}$$

where v_0 represents the value of the mean anomaly at an initial time (epoch), $t = \bar{t}_0$. Constant v_0 may be used as the sixth orbital element as well. The mean anomaly v_M is frequently used as an independent variable instead of time t, since a linear relation exists between the two quantities. Note that the mean anomaly is usually denoted by M in the astronomical literature; we have used a different notation since we reserve symbol M for the mass of the Earth.

The motion of a satellite satisfies Kepler's laws only if the gravitational field, in which the satellite moves, is spherically symmetrical. If the force field deviates from spherical symmetry, the orbital elements of the satellite are not constant, but vary with time. We speak of "perturbations" of the orbit due to a "perturbing force". In this case, there is a system of orbital elements at each instant of time which characterises the orbit at this very moment. This system of elements determines the elliptical motion of the satellite thereafter provided the perturbing force disappears at this moment. Such elements are called the *osculating orbital elements.*

12.3.2 Some relations between orbital elements and spherical coordinates. Relations between angles

We have arrived at the conclusion that it is convenient to describe satellite motion in terms of orbital elements. However, the gravitational field of the Earth, in which the satellite moves, has been described in spherical coordinates. Consequently, the relations between these two types of coordinates are required in solving many problems. Several of these relations will be derived here.

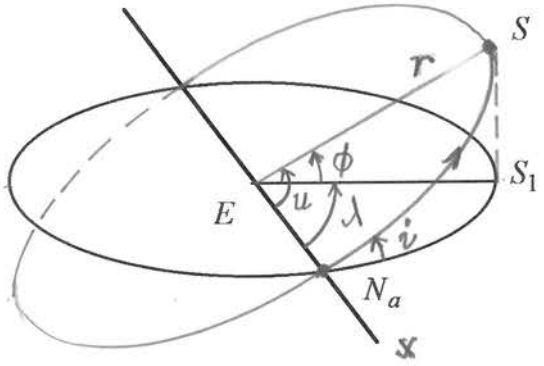

Figure 12.3: Spherical coordinates and orbital elements.

Introduce a Cartesian coordinate system whose origin is at the Earth's centre, E, the x- and y-axes lie in the equatorial plane, the x-axis coinciding with the nodal line, and the z-axis coinciding with the Earth's rotational axis (Fig. 12.3). Denote the corresponding spherical coordinates of the satellite by r, ϕ, λ. Note that in this case, ϕ is the geocentric latitude, but λ differs from the geographical longitude, as it is measured from the nodal line of the satellite.

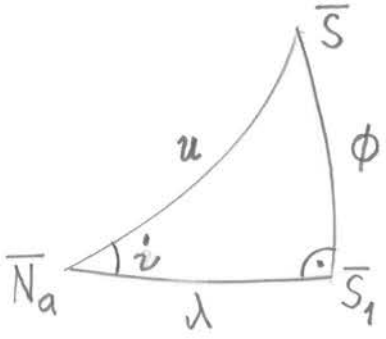

Figure 12.4: Projection of spherical triangle SS_1N_a in Fig. 12.3 onto a unit sphere with centre E. The projections of the vertices are denotes by primes.

Denote by S_1 the projection of satellite S onto the equatorial plane, and consider spherical triangle SS_1N_a, or its projection onto the unit sphere as shown in Fig. 12.4. It follows from the sine and cosine rules (see formula (5.37)) that

$$\frac{\sin\phi}{\sin i} = \frac{\sin u}{\sin 90^\circ} ,$$

$$\cos u = \cos\phi \cos\lambda + \sin\phi \sin\lambda \cos 90^\circ .$$

Consequently,

$$\sin\phi = \sin i \sin u , \tag{12.4}$$

$$\cos u = \cos\phi \cos\lambda . \tag{12.5}$$

The cosine rule also yields

$$\cos\phi = \cos u \cos\lambda + \sin u \sin\lambda \cos i .$$

Eliminate variable u in this relation by using (12.5) and (12.4):

$$\cos\phi = \cos\phi \cos^2\lambda + \frac{\sin\phi}{\sin i} \sin\lambda \cos i .$$

By dividing this equation by $\cos\phi$, we get

$$1 - \cos^2\lambda = \tan\phi \sin\lambda \cot i ,$$

which yields another relation,

$$\tan\phi = \tan i \sin\lambda . \tag{12.6}$$

12.3.3 Polar form of the equation of an ellipse

We have already derived several relations between spherical coordinates ϕ, λ, and orbital elements. Now, we shall express radial coordinate r in terms of orbital elements.

Introduce a Cartesian coordinate system in Fig. 12.2 in the following way: the coordinate origin coincides with the centre of the ellipse, C; one axis, denoted by x', passes through focus E; the second axis is denoted by y'. The equation of the ellipse can then be expressed in the well-known form

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1 , \tag{12.7}$$

where a and b are the semi-major and semi-minor axes, respectively.

Let us express the coordinates x' and y' of point S in terms of radial distance $r = \overline{ES}$ and true anomaly v (the origin of these polar coordinates is at the Earth' centre E). Denote by e^* the distance of the focus from the centre of the ellipse, $e^* = \overline{CE} = \sqrt{a^2 - b^2}$. It then holds that

$$x' = e^* + r\cos v \ ,$$

$$y' = r\sin v \ .$$

From the definition of the eccentricity,

$$e = e^*/a \ ,$$

we get

$$b^2 = a^2 - e^{*2} = a^2\left(1 - e^2\right) \ . \tag{12.8}$$

Inserting these expressions into Eq. (12.7) yields

$$e^2 + 2e\frac{r}{a}\cos v + \left(\frac{r}{a}\right)^2\left(\cos^2 v + \frac{\sin^2 v}{1-e^2}\right) = 1 \ .$$

We have obtain a quadratic equation in the unknown $\left(r/a\right)$,

$$\left(\frac{r}{a}\right)^2 \frac{1 - e^2\cos^2 v}{1-e^2} + 2\frac{r}{a}e\cos v - \left(1 - e^2\right) = 0 \ .$$

Its solution yields the so-called *polar equation of an ellipse (focal equation of an ellipse)* in the form

$$r = \frac{a\left(1-e^2\right)}{1 + e\cos v} \ . \tag{12.9}$$

This equation of an ellipse is also frequently expressed as

$$r = \frac{p}{1 + e\cos v} \ , \tag{12.10}$$

where $p = b^2/a$.

Note that the second solution of the quadratic equation yields negative values of r, i.e. the values along the half-line which is opposite to ES .

12.4 Satellite Motion in a Central Field. Circular Orbits

12.4.1 Central gravitational field

Hereafter in this chapter we consider a satellite to be a mass point whose mass m is negligible in comparison with the mass of the Earth, M, i.e., we assume

$$m \ll M \ . \tag{12.11}$$

This assumption will simplify the equations of satellite motion, since the Earth-satellite system need not be considered as a two-body system. In this case we may neglect the motion of the Earth about the common centre of mass, and consider the Earth to be fixed in an inertial reference frame. Moreover, we could put $m = 1$, since mass m will cancel out in the equations of motion.

In this chapter we shall revert to the signs of potentials as are common in physics, i.e. we shall use the potentials with signs opposite to those in Chapter 7. Hence, the gravitational potential of a mass point will be negative, $V < 0$, and the corresponding intensity will be $\mathbf{E} = -\operatorname{grad} V$.

In this section we assume the Earth (or another central body) to be spherically symmetrical. Its external gravitational potential is then

$$\boxed{V = -GM/r}, \tag{12.12}$$

where r is the radial distance from the Earth's centre. The components of intensity in spherical coordinates then read

$$E_r = -\frac{\partial V}{\partial r} = -\frac{GM}{r^2}, \quad E_\phi = E_\lambda = 0 \ . \tag{12.13}$$

Thus, only the radial component of intensity is non-zero in this field.

This spherically symmetrical gravitational field is a special case of a central field. Let us remind the reader that we speak of a *central field* if intensity $\mathbf{E}$ depends only on the distance from a fixed point, and is parallel to the corresponding radius vector, i.e.

$$\mathbf{E} = \mathbf{E}(r), \quad \mathbf{E} \parallel \mathbf{r} \ . \tag{12.14}$$

We shall prove below that the motion of a particle in any central field satisfies Kepler's second law. Moreover, if the force is attractive and decreases as $1/r^2$, also Kepler's first and third laws hold true. (The reader who knows the derivation of Kepler's laws may skip the corresponding parts of this chapter).

12.4.2 Kepler's second law

The torque of intensity **E** about the origin is defined as

$$\mathbf{N} = \mathbf{r} \times \mathbf{E} \ , \tag{12.15}$$

where **r** is the radius vector of the point of application of **E**.

Now, consider a central field and assume the coordinate origin to be at the centre of the field. Torque (12.15) is then equal to zero, $\mathbf{N} = 0$, because intensity **E** is parallel to radius vector **r**. Consequently, according to the angular momentum theorem (3.23), the angular momentum of the particle, **L**, is conserved in such a field:

$$\mathbf{L} = \mathbf{r} \times m\mathbf{v} = \mathbf{const} \ , \tag{12.16}$$

where **v** is the orbital velocity of the particle. Since vector **L** is perpendicular to the orbital plane, it follows from Eq. (12.16) that the orbital plane is kept fixed in an inertial system. In other words, the trajectory of a particle in a central field lies in a plane.

The magnitude of vector **L** is

$$L = mrv \sin\alpha = \text{const} \ , \tag{12.17}$$

where α is the angle between **r** and **v** (the letter v denotes the velocity here, not the true anomaly). Since the areal velocity is

$$S = \frac{1}{2} rv \sin\alpha \ , \tag{12.18}$$

equation (12.17) represents Kepler's second law for motion in a central field.

The derivation of the remaining two Kepler's laws is rather complicated if Newtonian mechanics is used. Therefore, let us first study the special case of circular orbits.

12.4.3 Circular orbit in a central field

Consider a satellite in a circular orbit at distance r from the Earth's centre. In this case, the gravitational force exerted on the satellite is balanced by the centrifugal force:

$$\frac{GMm}{r^2} = m\omega^2 r \ , \tag{12.19}$$

where ω is the angular velocity of the satellite motion. Equation (12.19) can be expressed as

$$\boxed{\omega^2 r^3 = GM}\,, \tag{12.20}$$

or

$$\frac{r^3}{T^2} = \frac{GM}{4\pi^2}\,, \tag{12.21}$$

where T is the period of revolution. We have derived Kepler's third law for circular orbits. Note that Kepler's third law for the two-body problem (but also for circular orbits only) had already been derived in Section 4.6; see formula (4.20), which represents a generalisation of (12.20).

By inserting orbital velocity $v = \omega r$ (progressive velocity) into Eq. (12.20), we immediately arrive at the well-known formula for the velocity of a satellite in a circular orbit:

$$v = \sqrt{\frac{GM}{r}}\,. \tag{12.22}$$

Formula (12.21) can be used for an approximate determination of the geocentric gravitational constant GM if distance r and orbital period T are measured. For a more accurate determination, however, various perturbations of the motion must be taken into account, especially if the orbit is close to the Earth. Consequently, contemporary determinations of GM are usually based on observations of distant satellites or on lunar laser ranging; for details we refer the reader to (Burša and Pec, 1993).

The parameters of three circular orbits, computed by means of (12.20), are given in Tab. 12.1. These orbits are as follows: a) the orbit close to the Earth's equator (the atmosphere is ignored); b) the orbit at a height of 300 km above the equator; c) the orbit of a stationary satellite, whose angular velocity ω is equal to the angular velocity of the Earth's rotation ω_E.

Table 12.1. Parameters of circular orbits of Earth's satellites: r is the distance from the Earth's centre, $h = r - a_0$ is the height above the equator, ω is the angular velocity of the satellite, $T = 2\pi/\omega$ is the orbital period, and $v = \omega r$ is the progressive velocity of the satellite. The corresponding parameters of the Earth are given in the text.

r(km)	h(km)	$\omega\left(\text{rad s}^{-1}\right)$	T	$v\left(\text{km s}^{-1}\right)$
6 378	0	1.239×10^{-3}	84.5^{m}	7.91
6 678	300	1.157×10^{-3}	90.5^{m}	7.73
42 164	35 786	7.292×10^{-5}	$23^{\text{h}}56.1^{\text{m}}$	3.07

For the computation of Tab. 12.1 we have used the following numerical values: $GM = 398\,600.44 \times 10^9\ \text{m}^3\text{s}^{-2}$, $a_0 = 6378\,\text{km}$ for the equatorial radius, and $\omega_E = 7.292\,115 \times 10^{-5}\ \text{rad s}^{-1}$; more accurate values can be found

in Section 1.11. Note that the equatorial radius of the Earth is denoted here by a_0, whereas symbol a is reserved for the semi-major axis of an orbit.

Note that formula (12.20) is only approximate, since the influence of the Earth's flattening and other terms are not considered there. If these terms are taken into account, some values in Tab. 12.1 would change slightly. For example, more accurate values for the orbit close to the equator can be obtained if GM in Eq. (12.20) is multiplied by the factor $\left(1-3J_2\right)$, where J_2 is the second-degree zonal geopotential parameter; see the derivation below. If value $J_2 = 1\,082.64 \times 10^{-6}$ is used, the parameters in the first line of Tab. 12.1 would be as follows: $\omega = 1.237 \times 10^{-3}$ rad s^{-1}, $T = 84.6$ minutes, and $v = 7.89$ km s^{-1}.

12.5 The Effect of the Earth's Flattening on a Circular Orbit

Let us return to formula (7.27) for the gravity potential of the Earth. Changing the sign, we can express the first two terms of the gravitational potential as

$$V = -\frac{GM}{r} - \frac{G(C-A)}{2r^3}\left(1-3\sin^2\phi\right), \tag{12.23}$$

where r is the radial distance from the Earth's centre, ϕ the geocentric latitude, C and A are the polar and equatorial moments of inertia, respectively. In the theory of satellite motion, potential (12.23) is frequently expressed as

$$V = -\frac{GM}{r}\left[1-\left(\frac{a_0}{r}\right)^2 J_2\, P_2\left(\sin\phi\right)\right], \tag{12.24}$$

where a_0 is the equatorial radius of the Earth, $P_2\left(\sin\phi\right) = \frac{3}{2}\sin^2\phi - \frac{1}{2}$ is the Legendre polynomial of the second degree, and

$$J_2 = \frac{C-A}{Ma_0^2} \tag{12.25}$$

is the second-degree zonal geopotential (Stokes') parameter. For the sake of brevity, we shall express potential (12.23) also as

$$V = -\frac{GM}{r} - \frac{K}{2r^3}\left(1-3\sin^2\phi\right), \tag{12.26}$$

where

$$K = G(C - A) = GMJ_0 a_0^2 \ . \tag{12.27}$$

12.5.1 The first term of the perturbing field

In the previous section we calculated the circular orbits in the central field described by potential $\left(-GM/r\right)$.

Now, let us study the main perturbations of a circular orbit due to the flattening of the Earth. In fact, we shall study only perturbations due to the second term of the potential series (the first term of the perturbing potential), which is closely related to the flattening. This potential, which we shall denote by V_2, is represented by the second term in (12.26), i.e.

$$\boxed{V_2 = -\frac{K}{2r^3}\left(1 - 3\sin^2\phi\right)} \ . \tag{12.28}$$

The corresponding intensity in spherical coordinates reads

$$\mathbf{E}_2 = -\operatorname{grad} V_2 = \left(E_r, E_\phi, E_\lambda\right) \ , \tag{12.29}$$

where

$$\begin{aligned} E_r &= -\frac{\partial V_2}{\partial r} = -\frac{3K}{2r^4}\left(1 - 3\sin^2\phi\right) = 3GM\frac{a_0^2}{r^4} J_2 P_2\left(\sin\phi\right) \ , \\ E_\phi &= -\frac{1}{r}\frac{\partial V_2}{\partial \phi} = -\frac{3K}{r^4}\sin\phi\cos\phi \ , \quad E_\lambda = -\frac{1}{r\cos\phi}\frac{\partial V_2}{\partial \lambda} = 0 \ . \end{aligned} \tag{12.30}$$

We shall also need the torque $\mathbf{N}_2$ of intensity $\mathbf{E}_2$ about the coordinate origin,

$$\mathbf{N}_2 = \mathbf{r} \times \mathbf{E}_2 \ , \tag{12.31}$$

where $\mathbf{r}$ is the radius vector of the satellite. The torque corresponding to component E_r is again equal to zero, so that the non-zero torque comes only from component E_ϕ. The latter torque lies in the equatorial plane, and is perpendicular to ES_1; see Fig 12.3. Its magnitude is (we omit subscript 2)

$$N = rE_\phi = -\frac{3K}{r^3}\sin\phi\cos\phi \ , \tag{12.32}$$

and its components are

$$N_x = N\sin\lambda \ , \quad N_y = -N\cos\lambda \ ; \tag{12.33}$$

see Fig. 12.5. Let us repeat that the x- and y-axes lie in the equatorial plane, and the x-axis coincides with the nodal line.

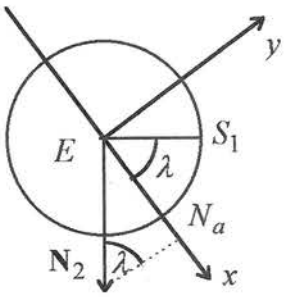

Figure 12.5: Decomposition of torque $\mathbf{N}_2$ in the equatorial plane.

12.5.2 Perturbations of a circular orbit

Intensity $\mathbf{E}_2$, which represents a deviation from the central field, causes *perturbations* of the orbit, i.e. deviations of the satellite motion from Kepler's laws. This intensity is generally different at different points of the orbit. As a result, orbital parameters are no longer constant, but vary with time in a rather complicated manner. However, we are not interested here in all the detailed variations, but only in certain averaged motions. For this purpose we shall calculate the mean values of intensity $\mathbf{E}_2$ and torque $\mathbf{N}_2$, and study their effects. In other words, we shall study only the long-term changes of the orbital elements.

First, consider radial intensity E_r. By inserting $\sin\phi = \sin i \sin u$ from (12.4), we get

$$E_r = -\frac{3K}{2r^4}\left(1 - 3\sin^2 i \, \sin^2 u\right) .$$

In a circular orbit, the argument of the latitude, u, varies uniformly with time. Thus, the mean value of $\sin^2 u$, averaged over one revolution, is equal to one half. The mean value of the radial intensity then reads

$$\overline{E_r} = -\frac{3K}{2r^4}\left(1 - \frac{3}{2}\sin^2 i\right) . \tag{12.34}$$

This term represents a small contribution to the main radial term $\left(-GM/r^2\right)$. The value of $\overline{E_r}$ depends on inclination i, and decreases rapidly with increasing distance from the Earth's centre, r. In order to separate the two radial terms, highly precise measurements of radii and periods for different orbits would be needed. Constant K could then be determined.

A more interesting effect is caused by the transverse component of intensity, E_ϕ. The component N_x of the corresponding torque can be expressed as

$$N_x = -\frac{3K}{r^3}\sin\phi\cos\phi\sin\lambda = -\frac{3K}{r^3}\sin\phi\cos\phi\frac{\tan\phi}{\tan i} =$$

$$= -\frac{3K}{r^3}\frac{\sin^2\phi}{\tan i} = -\frac{3K}{r^3}\sin i\cos i\sin^2 u \ ,$$

where we have used (12.33), (12.32), (12.6) and (12.4). Therefore, the mean value of this component, averaged over a period, is

$$\overline{N_x} = -\frac{3K}{2r^3}\sin i\cos i \ . \tag{12.35}$$

Component N_y and its mean value $\overline{N_y}$ can be expressed similarly. It follows from (12.33) and (12.32), if $\cos\lambda$ is expressed from (12.5) and $\sin\phi$ from (12.4), that

$$N_y = \frac{3K}{2r^3}\sin i\sin 2u \ .$$

As the mean value of $\sin 2u$ is zero, we get $\overline{N_y} = 0$. Therefore, component N_y has no long-term effect.

Since the angular momentum of the satellite is perpendicular to the orbital plane, torque N_x causes a regression of the nodal line, whereas torque N_y causes changes of the inclination. Let us consider only the regression of the nodal line.

Note that torques N_x and N_y, introduced above, relate to the intensity. Thus the torques of the force exerted on a satellite of mass m are mN_x and mN_y.

Figure 12.6 shows the relation between the change of the satellite's angular momentum, ΔL, and the corresponding change of the longitude of the ascending node, $\Delta\Omega$:

$$\Delta L = L\sin i\,\Delta\Omega \ . \tag{12.36}$$

Divide this equation by the time interval Δt, during which these changes occurred. Taking the limit for $\Delta t \to 0$, and using the angular momentum theorem (3.23) yields

$$\frac{\mathrm{d}L}{\mathrm{d}t} = L\sin i\frac{\mathrm{d}\Omega}{\mathrm{d}t} = mN_x \ . \tag{12.37}$$

In a circular orbit, the angular momentum of the satellite is $L = mr^2\omega$, where r is the radius of the orbit and ω the angular velocity of the satellite. After averaging Eq. (12.37) and using (12.35), we get

$$\dot{\Omega}_{mean} = \left(\frac{\mathrm{d}\Omega}{\mathrm{d}t}\right)_{mean} = \frac{1}{r^2\omega}\left(-\frac{3K}{2r^3}\right)\cos i \ . \tag{12.38}$$

By substituting for K from its definition (12.27) and for GM from Kepler's third law, i.e.

$$K = GMJ_2a_0^2 = \omega^2 r^3 J_2 a_0^2 \ ,$$

we arrive at the final result in the form

$$\boxed{\dot{\Omega}_{mean} = -\frac{3}{2}\omega J_2 \frac{a_0^2}{r^2}\cos i} \ . \tag{12.39}$$

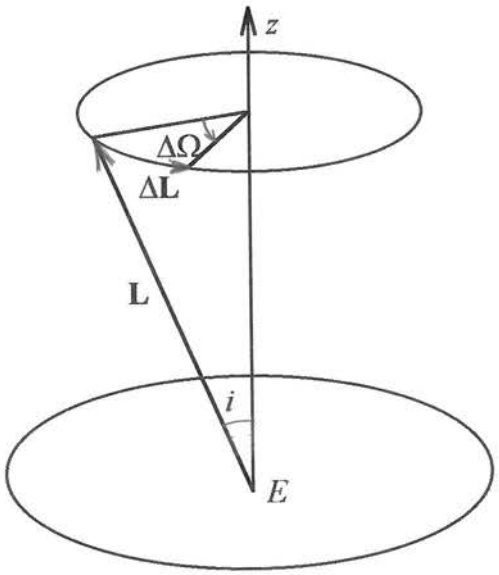

Figure 12.6: Relation between the change $\Delta\mathbf{L}$ of the angular momentum, and the corresponding change $\Delta\Omega$ of the longitude of the ascending node.

For a circular orbit, this formula gives the mean change of the longitude of the ascending node (regression of the nodal line) due to the Earth's flattening. The flattening is represented here by coefficient J_2.

In order to utilise the Earth's rotation from west to east, satellites are usually launched eastwards, so that the inclinations of orbits are acute angles, and $\cos i > 0$; see Fig. 12.3. The negative sign in (12.39) then means that the nodal

line of such an orbit moves westwards, i.e. ascending node N_a will be shifted in the next revolution westwards relative to its previous position (Fig. 12.7). This means that, as opposed to the motion in a central field, the trajectory of the satellite is not now generally a plane curve.

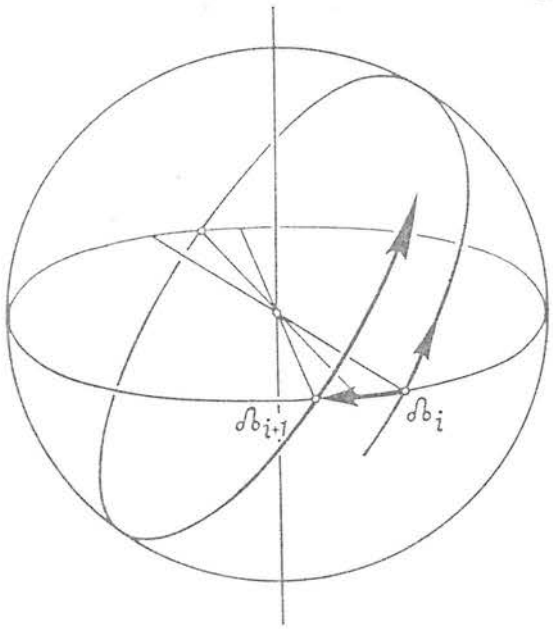

Figure 12.7: Regression of the nodal line of the satellite's orbital plane. (After (Burša and Pec, 1993)).

Let us remind the reader that the motion, just described, refers to an inertial system. As the Earth rotates below the orbit, all observations of satellites from the Earth's surface must be corrected for the Earth's rotation. However, this is only a small complication in studying satellite orbits.

The regression of the nodal line can easily be explained, even on the basis of an elementary consideration, in the following way. Consider a satellite which has passed through the ascending node and moves above the northern hemisphere in a north-eastern direction (Fig. 12.7). The distribution of masses of the Earth's equatorial bulge is asymmetrical with respect to the orbital plane. Consequently, the nearer masses of the bulge (from the equatorial region on the right-hand side of Fig. 12.7) turn the motion of the satellite to the "right". Consequently, the next intersection of the orbit with the equatorial plane will be shifted westwards relative to the unperturbed motion.

Having measured the parameters (osculating parameters) of the circular orbit and the regression of its node, we can use formula (12.39) to determine the coefficient J_2 and afterwards the Earth's flattening. The first satellite determinations of the Earth's flattening followed this very procedure (Buchar, 1958).

To demonstrate the magnitude of the regression of the nodal line, let us consider circular orbits at different distances from Earth's centre, but with the same inclination, say, $i = 30°$. Table 12.2 contains the results for the orbits which have already been considered in Tab. 12.1. We have used the following constants: $a_0 = 6\,378\,137\,\text{m}$ for the equatorial radius, and $J_2 = 1\,082.64 \times 10^{-6}$ for the second-order zonal geopotential parameter. The last column of the table gives the regression in degrees per solar day.

Table 12.2 indicates that the motion of the nodal line may be as much as *several degrees per day* for satellites which are close to the Earth. This is a large value in comparison with the many small perturbations which are measured in astronomy (e.g., the Earth's precession). Hence, the regression of the nodal line can easily be measured with a high accuracy by astronomical methods.

On the other hand, the velocity of the motion of the node decreases rapidly with increasing distance from the Earth. For example, for a satellite at a distance of the Moon, i.e. $r = 384\,400\,\text{km}$, $\omega = \omega_L = 2.6619 \times 10^{-6}\ \text{rad s}^{-1}$, inclination again $i = 30°$, formula (12.39) yields (in the absolute value) only $\dot{\Omega}_{mean} = 1.03 \times 10^{-12}\ \text{rad s}^{-1}$, which is only 0.2° per century. Although, in principle, the Earth's flattening could also have been studied in the pre-satellite era astronomically from the motion of the Moon, the above estimate indicates the futility of such attempts. This also demonstrates the revolutionary contribution of satellites to studies of the large-scale properties of the Earth's gravitational field.

Table 12.2. Regressions of the nodal line for circular orbits of inclination $i = 30°$. Notation: r is the distance from the Earth's centre, $h = r - a_0$ is the height above the equator, $\dot{\Omega}$ is the mean regression of the nodal line. The other parameters of the orbits are given in Tab.12.1, and the parameters of the Earth are given in the text.

r(km)	h(km)	$\dot{\Omega}\left(\text{rad s}^{-1}\right)$	$\dot{\Omega}\left(\text{degrees per day}\right)$
6 378	0	-1.74×10^{-6}	−8.63
6 678	300	-1.48×10^{-6}	−7.35
42 164	35 786	-2.35×10^{-9}	−0.01

12.6 Relations between Precession and Perturbations of a Circular Orbit

We have learnt in Chapter 4 that the gravitational effect of the Moon and Sun on the flattened Earth causes the precession of its rotational axis. Now we have studied an opposite phenomenon, namely the effect of the flattened Earth on another body. It is evident that there is a relationship between these two effects,

since both are connected with the gravitational torque due to the flattened Earth.

The flattened Earth exerts torque $m\mathbf{N}_2$ on a satellite of mass m; see formulae (12.31) to (12.33). According to the angular momentum theorem, this torque is equal to the time derivative of the angular momentum of the satellite, $\mathrm{d}\mathbf{L}/\mathrm{d}t = m\mathbf{N}_2$. On the other hand, according to Newton's third law, the satellite exerts torque $\left(-m\mathbf{N}_2\right)$ on the Earth. And again according to the angular momentum theorem, this torque is equal to the time derivative of the angular momentum of the Earth:

$$\frac{\mathrm{d}\mathbf{L}_E}{\mathrm{d}t} = -m\mathbf{N}_2 \ . \tag{12.40}$$

The only small difference is that the angular momentum of the satellite comes from its orbital motion, $L = mr^2\omega$, whereas the angular momentum of the Earth in connected with its rotation, i.e.

$$L_E = C\omega_E \ , \tag{12.41}$$

where C is the polar moment of inertia and ω_E is the angular velocity of the Earth's rotation.

For the precession we can draw a figure similar to Fig. 12.6, but the role of the axes in the figure must be interchanged. Vector $\mathbf{L}$, being replaced by vector $\mathbf{L}_E$, would determine the direction of the Earth's rotation axis, whereas axis z would be perpendicular to the orbital plane of the perturbing body. Consequently, instead of (12.37) we now get

$$\frac{\mathrm{d}L_E}{\mathrm{d}t} = \omega_p L_E \sin i = -mN_x \ , \tag{12.42}$$

where ω_p is the angular velocity of precession; this replaces $\mathrm{d}\Omega/\mathrm{d}t$ in (12.37). Insert the mean values into Eq. (12.42), and express $\overline{N_x}$ from (12.35) and L_E from (12.41). Hence,

$$\overline{\omega_p} = \frac{3Km}{2r^3 C\omega_E}\cos i \ . \tag{12.43}$$

Since $K = G(C - A)$, see (12.27), we arrive at

$$\boxed{\overline{\omega_p} = \frac{3}{2}\frac{G}{\omega_E}\frac{C-A}{C}\frac{m}{r^3}\cos i} \ . \tag{12.44}$$

This formula expresses the *angular velocity of precession* of the Earth's rotational axis due to a satellite of mass m which revolves in a circular orbit of radius r and inclination i.

If the mass of the satellite is small (an artificial satellite), the corresponding precession of the Earth's axis is quite negligible. Only a sufficiently massive body (the Moon, the Sun) can generate a measurable precession.

Formula (12.44) agrees with formulae (4.1) and (4.2) for the precession due to the Sun and Moon, where only different notations are used. Hence, we have finally derived the formulae for the luni-solar precession, which were presented in Chapter 4 without proof.

12.7 Elliptical Orbits in a Central Field

12.7.1 Kepler's problem. Derivation of the equations of motion

We shall now consider Kepler's problem, i.e. the determination of the orbits of two particles interacting according to Newton's gravitational law. By introducing a reduced mass, this two-body problem with central forces can always be reduced to a one-body problem (Kittel et al., 1965). By the one-body problem we understand the motion of a particle in a central field which is fixed in an inertial system.

If the mass M of one particle (planet) is much greater than the mass m of the other particle (artificial satellite), the motion of the massive particle about the common centre of mass may be neglected. In this case we immediately arrive at a one-body problem. Here we shall restrict ourselves to this problem.

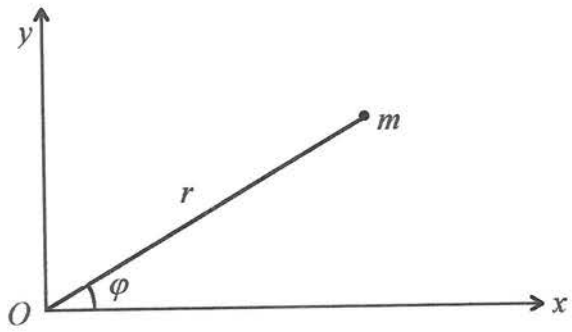

Figure 12.8: Coordinates used in solving Kepler's problem.

Therefore, consider a particle of mass M which is fixed at the origin of an inertial system. The potential of its gravitational field is

$$V^* = -\frac{GM}{r} , \tag{12.45}$$

where r is the distance from the coordinate origin, O, see Fig. 12.8. Consider another particle, mass m, moving in this fixed field. The potential energy of this particle in such a field is

$$V = mV^* = -\frac{GMm}{r} . \tag{12.46}$$

Since the force field is central, particle m will move in one plane, say the (x, y)-plane. However, instead of the Cartesian coordinates x and y of the particle, we shall use polar coordinates, r, φ:

$$x = r\cos\varphi , \quad y = r\sin\varphi . \tag{12.47}$$

Polar coordinates will be more convenient for solving our problem. The components of the particle velocity are

$$\dot{x} = \dot{r}\cos\varphi - r\sin\varphi\,\dot{\varphi} , \quad \dot{y} = \dot{r}\sin\varphi + r\cos\varphi\,\dot{\varphi} , \tag{12.48}$$

and the corresponding accelerations read

$$\begin{aligned} \ddot{x} &= \ddot{r}\cos\varphi - 2\dot{r}\sin\varphi\,\dot{\varphi} - r\cos\varphi\,\dot{\varphi}^2 - r\sin\varphi\,\ddot{\varphi} , \\ \ddot{y} &= \ddot{r}\sin\varphi + 2\dot{r}\cos\varphi\,\dot{\varphi} - r\sin\varphi\,\dot{\varphi}^2 + r\cos\varphi\,\ddot{\varphi} . \end{aligned} \tag{12.49}$$

Since the force acting on the particle is $\mathbf{F} = -\mathrm{grad}V$, the equations of motion of the particle in the Cartesian system are as follows:

$$-\frac{\partial V}{\partial x} = m\ddot{x} , \quad -\frac{\partial V}{\partial y} = m\ddot{y} . \tag{12.50}$$

Instead of these equations of motion we are seeking the equations of motion in the direction of radial coordinate r and in the direction perpendicular to it.

The force acting in the radial direction is

$$-\frac{\partial V}{\partial r} = -\frac{\partial V}{\partial x}\frac{\partial x}{\partial r} - \frac{\partial V}{\partial y}\frac{\partial y}{\partial r} = -\frac{\partial V}{\partial x}\cos\varphi - \frac{\partial V}{\partial y}\sin\varphi .$$

Insert into the left-hand side of this equation from (12.46), and into the right-hand side from (12.50) and (12.49). After some simple algebra, we arrive at the following equation of motion:

$$\boxed{\ddot{r} - r\dot{\varphi}^2 + \frac{GM}{r^2} = 0}\ . \tag{12.51}$$

The force in the perpendicular direction (along the coordinate axis φ) is

$$-\frac{1}{r}\frac{\partial V}{\partial \varphi} = -\frac{1}{r}\left(\frac{\partial V}{\partial x}\frac{\partial x}{\partial \varphi} + \frac{\partial V}{\partial y}\frac{\partial y}{\partial \varphi}\right) = \frac{\partial V}{\partial x}\sin\varphi - \frac{\partial V}{\partial y}\cos\varphi\ .$$

Since $\partial V/\partial\varphi = 0$, we obtain the simple equation,

$$2\dot{r}\dot{\varphi} + r\ddot{\varphi} = 0\ .$$

After multiplying this equation by r, the new equation may be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\dot{\varphi}\right) = 0\ .$$

Consequently,

$$\boxed{r^2\dot{\varphi} = C}\ , \tag{12.52}$$

where C is a constant. The last equation represents Kepler's second law, since $r^2\dot{\varphi}$ is double the areal velocity; see formulae (12.17) and (12.18).

Equations (12.51) and (12.52) are the required equations of motion for a particle in the central field, described by potential (12.45). Their solution will be given further on in this chapter.

12.7.2 Modifications of the equations of motion. Binet's formula

We shall solve Eqs. (12.51) and (12.52) for the motion of a particle in the central gravitational field with potential (12.45). These equations represent a system of two coupled non-linear differential equations in $r = r(t)$ and $\varphi = \varphi(t)$. Although $\dot{\varphi}$ in (12.51) may be eliminated by means of (12.52), the equation for $r(t)$ remains very complicated:

$$\ddot{r} - \frac{C^2}{r^3} + \frac{GM}{r^2} = 0\ . \tag{12.53}$$

We do not know how to solve this equation analytically, in order to determine $r = r(t)$.

However, there is another possibility. The law of areas, i.e. Eq. (12.52), determines the velocity of the particle if its trajectory is known. Therefore, we

shall first attempt to determine the shape of the trajectory. For this reason, solve Eq. (12.53) for r as a function $r(\varphi)$ of angle $\varphi = \varphi(t)$. This yields

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\varphi}\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\varphi}\frac{C}{r^2}\,, \tag{12.54}$$

using (12.52). Then, upon differentiating once more,

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 r}{\mathrm{d}\varphi^2}\frac{C^2}{r^4} + \frac{\mathrm{d}r}{\mathrm{d}\varphi}\left(\frac{-2C}{r^3}\right)\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}^2 r}{\mathrm{d}\varphi^2}\frac{C^2}{r^4} - \frac{2C^2}{r^5}\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2\,, \tag{12.55}$$

using (12.52) again, and also (12.54). Equation (12.53) now becomes

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\varphi^2}\frac{C^2}{r^4} - \frac{2C^2}{r^5}\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2 - \frac{C^2}{r^3} + \frac{GM}{r^2} = 0\,. \tag{12.56}$$

Equation (12.56) has an even more complicated form than the original equation (12.53). Nevertheless, these equations can be simplified by substituting

$$\boxed{w(\varphi) = \frac{1}{r(\varphi)}}\,. \tag{12.57}$$

The derivatives of this function are

$$\frac{\mathrm{d}w}{\mathrm{d}\varphi} = -\frac{1}{r^2}\frac{\mathrm{d}r}{\mathrm{d}\varphi}\,,\quad \frac{\mathrm{d}^2 w}{\mathrm{d}\varphi^2} = -\frac{1}{r^2}\frac{\mathrm{d}^2 r}{\mathrm{d}\varphi^2} + \frac{2}{r^3}\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2\,. \tag{12.58}$$

On comparing (12.58) with (12.55) we see that

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{C^2}{r^2}\frac{\mathrm{d}^2 w}{\mathrm{d}\varphi^2}\,, \tag{12.59}$$

so that the equation of motion (12.53) may now be expressed as

$$\boxed{\frac{\mathrm{d}^2 w}{\mathrm{d}\varphi^2} + w = \frac{GM}{C^2}}\,. \tag{12.60}$$

This important equation is known as *Binet's formula*. As opposed to Eqs. (12.53) or (12.56), Binet's formula is a linear differential equation, which can easily be solved (see below).

Let us add a historical remark which concerns Eq. (12.56). This equation is very complicated, so that it is not surprising that the suitable substitution (12.57) was found only after a hundred years of experience. However, on the other hand, already Kepler knew that the orbit was an ellipse. And in the polar equation (12.10), angle v appears in the denominator. Hence, the information from observations indicated that the reciprocal substitution, $1/r$, should have been tested first of all. In other words, experiments prompted the suitable substitution, but this information was ignored by theoreticians for a long time. This example from history demonstrates the importance of physical observations and measurements even in solving some purely theoretical problems.

12.7.3 Solution of the equations of motion. Derivation of Kepler's laws

Binet's formula (12.60) is a linear inhomogeneous differential equation of the second order. The corresponding homogeneous equation, i.e.

$$\frac{\mathrm{d}^2 w}{\mathrm{d}\varphi^2} + w = 0 \,, \tag{12.61}$$

resembles the well-known equation of the harmonic oscillator. Its general solution is $w = A\cos(\varphi + \alpha)$, where A and α are constants. If we add constant GM/C^2 to this solution, we have

$$w = A\cos(\varphi + \alpha) + \frac{GM}{C^2} \,, \tag{12.62}$$

which is the solution of (12.60).

By returning to the reciprocal function, $r = 1/w$, we obtain

$$r = \frac{p}{1 + e\cos v} \,, \tag{12.63}$$

where we have put $v = \varphi + \alpha$, $e = AC^2/(GM)$ and $p = C^2/(GM)$. Equation (12.63) is the polar form of the equation of a conic section, see (12.10). Constant e is the eccentricity and angle v is the true anomaly. The four types of possible curves described by (12.63) are:

circle	$e = 0$,
ellipse	$0 < e < 1$,
parabola	$e = 1$,
hyperbola	$e > 1$.

We have just demonstrated that closed orbits are elliptical, i.e. we have derived Kepler's first law.

We are now going to derive Kepler's third law. If $\mathrm{d}S$ is the area which the radius vector sweeps through in time $\mathrm{d}t$, then the areal velocity is

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{C}{2}\,, \tag{12.64}$$

see Eq. (12.52). Integrating this equation over one period T of the motion yields

$$S = \frac{CT}{2}\,. \tag{12.65}$$

Here S is the area of the ellipse, which can also be expressed as $S = \pi ab$, where a is the semi-major axis and b the semi-minor axis. Formula (12.65) then yields

$$2\pi ab = CT\,. \tag{12.66}$$

By comparing $p = b^2/a$ in (12.10) with $p = C^2/(GM)$ in (12.63) we find that

$$\frac{b^2}{a} = \frac{C^2}{GM}\,. \tag{12.67}$$

If the square of (12.66) is divided by (12.67), we arrive at

$$\boxed{\frac{a^3}{T^2} = \frac{GM}{4\pi^2}}\,. \tag{12.68}$$

This formula represents *Kepler's third law* for an elliptical orbit. It is a generalisation of formula (12.21) for a circular orbit.

Note the interesting fact that we had only two equations of motion at our disposal to derive three Kepler's laws. Equation (12.52) yielded Kepler's second law, and Eq. (12.51), in combination with (12.52), yielded Kepler's first law. However, the solutions of these equations contained integration constants which were not quite arbitrary, but certain relations existed between them, in particular $p = C^2/(GM)$. This relation then yielded Kepler's third law.

Chapter 13

Satellite Methods of Studying the Gravitational Field - Applications of Analytical Mechanics

13.1 Lagrange's Equations of the Second Kind

Newton's second law,

$$\mathbf{F} = m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\,, \tag{13.1}$$

represents three scalar equations of motion of a mass point (particle) in an inertial system. In a Cartesian coordinate system this is usually expressed as

$$F_i = m\frac{d^2x_i}{dt^2}\,,\quad i = 1, 2, 3\,. \tag{13.2}$$

We have seen that the form of the equation of motion changed in a non-inertial system, where fictitious forces had to be added to equation (13.1). Analogous reformulations of the equation of motion are needed also in inertial systems, if the character of the problem calls for the application of a special coordinate system. Studies of periodic orbits belong to this category of problems.

The direct transformation of the equations of motion in the Newtonian form (13.1) or (13.2) into a new coordinate system may be complicated. (Nevertheless, a simple case of such a transformation was described in Subsection 12.7.1). Therefore, a more general formulation of the equations of motion is desirable. These requirements are satisfied, for example, by Lagrange's equations of the second kind. They have the following form:

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} = 0}\,, \tag{13.3}$$

where q_j are generalised coordinates (various lengths, angles or other parameters), $\dot{q}_j = \mathrm{d}q_j/\mathrm{d}t$ are generalised velocities, $L = T - V$ is the Lagrangian function (the Lagrangian of the system), T is the kinetic energy and V the potential energy of the system, $j = 1,\ldots,n$, n being the number of degrees of freedom. The derivation of Lagrange's equations can be found in textbooks on analytical mechanics.

Note that a free particle has 3 degrees of freedom, i.e. $n = 3$. For a system of N free particles we have $n = 3N$. One particle whose motion is confined to a plane has 2 degrees of freedom, etc.

In the next section we shall use Lagrange's equations to derive the equations of motion in polar coordinates for a particle moving in a central gravitational field. An elementary derivation of these equations of motion was performed in Section 12.7.

13.2 Kepler's Problem. Derivation of the Equations of Motion by Means of Lagrange's Equations

In our problem, the kinetic energy of the particle is

$$T = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\varphi}^2\right) , \tag{13.4}$$

where we have used (12.48). Note that an experienced reader should be able to write the last expression for the kinetic energy at once, taking into account only the fact that the radial velocity is $v_r = \dot{r}$ and the perpendicular component is $v_\varphi = r\dot{\varphi}$.

The Lagrangian, $L = T - V$, reads

$$L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\varphi}^2\right) + \frac{GMm}{r} . \tag{13.5}$$

Putting $q_1 = r$, $q_2 = \varphi$, and inserting (13.5) into Lagrange's equations (13.3), we immediately arrive at the equations of motion:

$$m\ddot{r} - mr\dot{\varphi}^2 + \frac{GMm}{r^2} = 0 , \quad mr^2\dot{\varphi} = \text{const} . \tag{13.6}$$

These equations are identical with Eqs. (12.51) and (12.52). As opposed to their derivation in Section 12.7, the derivation presented here, using Lagrange's equations, was much simpler. For the solution of these equations we refer the reader to Section 12.7.

13.3 Solution of the Satellite Motion in a General Potential Field by Means of the Hamilton-Jacobi Equation

Perturbations of elliptical orbits represent a classical problem of celestial mechanics. Its solution is very complicated in terms of Newtonian mechanics or

Lagrange's equations. Consequently, the classical solutions of this problem are based on more advanced methods of analytical mechanics, in particular on Hamilton's equations and the Hamilton-Jacobi equation. It is assumed that the reader of this section has a basic knowledge of these methods.

The traditional courses of theoretical mechanics usually start with the Newtonian formulation of the equations of motion, and then pass through Lagrange's equations, Hamilton's equations to the Hamilton-Jacobi theory. All these methods are demonstrated on several traditional problems, such as the mathematical pendulum, the rolling of a cylinder on an inclined plane, and Kepler's problem. However, the drawback of these demonstrations is that the corresponding problems are relatively simple, so that the application of Lagrange's equations (or even the Newtonian mechanics) to their solution is quite sufficient. In the light of these examples, the Hamilton theory and the Hamilton-Jacobi theory look like unnecessary complications. Only the perturbations of orbits represent a really difficult problem, where the advantages and power of the Hamilton-Jacobi theory are fully appreciated. However, the complicated geometry and transformations of coordinates cause that this problem is not usually included in standard textbooks of theoretical mechanics. Nevertheless, it represents a classical, non-trivial application of the Hamilton-Jacobi theory.

After reviewing some basic relations, we shall solve the unperturbed problem, i.e. Kepler's problem, in order to determine the canonical constants. The time variations of these "constants" in the case of the perturbed motion are then briefly derived. Finally, the effect of the first term of the perturbing potential (actually, the effect of the Earth's flattening) on an elliptical orbit will be calculated in Section 13.4.

13.3.1 Basic relations

Let q_j, where $j = 1,2,\ldots,n$, be the generalised coordinates, and p_j the generalised momenta (impulses) of a mechanical system, defined by

$$p_j = \frac{\partial L}{\partial \dot{q}_j}, \tag{13.7}$$

where $L = T - V$ is the Lagrangian (see Section 13.1). The generalised coordinates and momenta are combined in Hamilton's canonical equations,

$$\dot{q}_j = \frac{\partial H}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial H}{\partial q_j}, \tag{13.8}$$

where $H = T + V$ is the Hamiltonian. This is, in general, a function of the type $H = H(q_j, p_j, t)$.

Very often we need to switch to new generalised coordinates, Q_j, and new generalised momenta, P_j, which preserve the form of Hamilton's equations:

$$\dot{Q}_j = \frac{\partial \overline{H}}{\partial P_j}, \quad \dot{P}_j = -\frac{\partial \overline{H}}{\partial Q_j}, \tag{13.9}$$

where $\overline{H}$ is a new Hamiltonian. This transition may be effected through a generating function, F. If the generating function is a function of the old generalised coordinates, new generalised coordinates, and of time, i.e. $F = F\left(q_j, Q_j, t\right)$, then the transformation relations take the form:

$$p_j = \frac{\partial F}{\partial q_j}, \quad P_j = -\frac{\partial F}{\partial Q_j}, \quad \overline{H} = H + \frac{\partial F}{\partial t}. \tag{13.10}$$

We shall seek a special generating function, denoted by S, to render the new Hamiltonian identically equal to zero, $\overline{H} \equiv 0$. According to (13.10), generating function S must satisfy the equation

$$\boxed{H\left(q_j, \frac{\partial S}{\partial q_j}, t\right) + \frac{\partial S}{\partial t} = 0}, \tag{13.11}$$

which is called the *Hamilton-Jacobi equation*. If some solution S of this equation is found, it follows from (13.9) that the new generalised coordinates and momenta are constant:

$$Q_j = \alpha_j, \quad P_j = \beta_j. \tag{13.12}$$

The set of constants α_j and β_j then describes the mechanical system in full. Formulae (13.10) now take the form:

$$p_j = \frac{\partial S}{\partial q_j}, \quad \beta_j = P_j = -\frac{\partial S}{\partial Q_j} = -\frac{\partial S}{\partial \alpha_j}. \tag{13.13}$$

An important category of problems is represented by the problem with a perturbing term. This problem is complicated, so that its Hamilton-Jacobi equation cannot be solved and the canonical constants determined, but is close enough to a simpler problem for which the canonical constants are known. Thus, assume the original Hamiltonian to have the form

$$H = H_0 + H_1 = H - R, \tag{13.14}$$

where H_0 is the unperturbed Hamiltonian (for which the canonical constants can be determined), H_1 is the perturbed Hamiltonian, and $R = -H_1$ is the so-called perturbing function. Consider a generating function, S_0, which satisfies the Hamilton-Jacobi equation for the unperturbed problem, i.e.

$$H_0 + \frac{\partial S_0}{\partial t} = 0 \ . \tag{13.15}$$

Apply the same generating function S_0 also to the perturbed problem. Using (13.10), (13.14) and (13.15), the new Hamiltonian is

$$K = H + \frac{\partial S_0}{\partial t} = H_0 + H_1 + \frac{\partial S_0}{\partial t} = H_1 = -R \ . \tag{13.16}$$

Consequently, for the new generalised coordinates α_j and momenta β_j we obtain from (13.9)

$$\dot{\alpha}_j = \frac{\partial K}{\partial \beta_j} = -\frac{\partial R}{\partial \beta_j}, \quad \dot{\beta}_j = -\frac{\partial K}{\partial \alpha_j} = \frac{\partial R}{\partial \alpha_j} \ . \tag{13.17}$$

As opposed to the unperturbed problem, the new generalised coordinates and momenta are not constant, but they vary with time.

13.3.2 Unperturbed solution

Consider a particle of mass $m = 1$, revolving about a fixed particle of mass M. Put $\mu = GM$, and express the Lagrangian of the moving particle in spherical coordinates r, ϕ, λ:

$$L = T - V = \frac{1}{2}\left[\dot{r}^2 + r^2\dot{\phi}^2 + \left(r^2\cos^2\phi\right)\dot{\lambda}^2\right] + \frac{\mu}{r} \ . \tag{13.18}$$

In order to obtain the Hamiltonian, we must replace the generalised velocities $\dot{q}_j$ by the generalised momenta using (13.7).

$$p_r = \frac{\partial L}{\partial \dot{r}} = \dot{r} \ , \quad p_\phi = \frac{\partial L}{\partial \dot{\phi}} = r^2\dot{\phi} \ , \quad p_\lambda = \frac{\partial L}{\partial \dot{\lambda}} = r^2\cos^2\phi\,\dot{\lambda} \ . \tag{13.19}$$

Hence

$$H = T + V = \frac{1}{2}\left[p_r^2 + \frac{p_\phi^2}{r^2} + \frac{1}{r^2 \cos^2\phi} p_\lambda^2 \right] - \frac{\mu}{r} \ . \qquad (13.20)$$

The Hamilton-Jacobi equation is then

$$\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \phi}\right)^2 + \frac{1}{r^2 \cos^2\phi}\left(\frac{\partial S}{\partial \lambda}\right)^2 - \frac{2\mu}{r} + 2\frac{\partial S}{\partial t} = 0 \ . \qquad (13.21)$$

We are seeking a special solution in the following form (we do not need any general solution, just one solution is sufficient):

$$S = S_r + S_\phi + S_\lambda - \alpha_1 t \ , \qquad (13.22)$$

where S_r is a function of r only, S_ϕ is a function of ϕ only, S_λ is a function of λ only, and α_1 is a constant. Inserting this trial solution into the Hamilton-Jacobi equation (13.21), we obtain

$$\left(\frac{\partial S_r}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S_\phi}{\partial \phi}\right)^2 + \frac{1}{r^2 \cos^2\phi}\left(\frac{\partial S_\lambda}{\partial \lambda}\right)^2 = \frac{2\mu}{r} + 2\alpha_1 \ . \qquad (13.23)$$

Only term $\partial S_\lambda / \partial \lambda$ depends on λ, so that we can put it equal to a constant, α_3. Multiplying Eq. (13.23) by r^2, we get

$$r^2\left(\frac{\partial S_r}{\partial r}\right)^2 - 2\alpha_1 r^2 - 2\mu r = -\left(\frac{\partial S_\phi}{\partial \phi}\right)^2 - \frac{\alpha_3^2}{\cos^2\phi} \ . \qquad (12.24)$$

The left-hand side of this equation depends on r, but the right-hand side does not. Thus we can put them equal to a constant, $-\alpha_2^2$. We then obtain three separate equations:

$$\frac{\partial S_\lambda}{\partial \lambda} = \alpha_3 \ , \quad \frac{\partial S_r}{\partial r} = \frac{L}{r} \ , \quad \frac{\partial S_\phi}{\partial \phi} = N \ , \qquad (13.25)$$

where we have introduced

$$L = \sqrt{2\alpha_1 r^2 + 2\mu r - \alpha_2^2} \ , \quad N = \sqrt{\alpha_2^2 - \frac{\alpha_3^2}{\cos^2\phi}} \ ; \qquad (13.26)$$

note that this L differs from the Lagrangian. Consequently, one solution of the Hamilton-Jacobi equation (13.21) may be expressed as

$$S = \int_{r_1}^{r} \frac{L}{r} \mathrm{d}r + \int_{0}^{\phi} N \, \mathrm{d}\phi + \alpha_3 \lambda - \alpha_1 t \ , \tag{13.27}$$

where the lower limits of the integrals may be chosen arbitrarily. The other three constants, β_j, are given by (13.13):

$$-\beta_1 = \frac{\partial S}{\partial \alpha_1} = \frac{\partial S_r}{\partial \alpha_1} - t \ \Rightarrow \ t - \beta_1 = \frac{\partial S_r}{\partial \alpha_1} = \int_{r_1}^{r} \frac{r}{L} \mathrm{d}r \ , \tag{13.28}$$

$$-\beta_2 = \frac{\partial S}{\partial \alpha_2} = \frac{\partial S_r}{\partial \alpha_2} + \frac{\partial S_\phi}{\partial \alpha_2} = -\int_{r_1}^{r} \frac{\alpha_2}{rL} \mathrm{d}r + \int_{0}^{\phi} \frac{\alpha_2}{N} \mathrm{d}\phi \ , \tag{13.29}$$

$$-\beta_3 = \frac{\partial S}{\partial \alpha_3} = \frac{\partial S_\phi}{\partial \alpha_3} + \lambda = \lambda - \int_{0}^{\phi} \frac{\alpha_3}{N \cos^2 \phi} \mathrm{d}\phi \ . \tag{13.30}$$

Relations (13.13) also yield the expressions for p_j:

$$p_r = \frac{\partial S}{\partial r} = \frac{\partial S_r}{\partial r} = \frac{L}{r} \ , \ p_\phi = \frac{\partial S_\phi}{\partial \phi} = N \ , \ p_\lambda = \frac{\partial S_\lambda}{\partial \lambda} = \alpha_3 \ . \tag{13.31}$$

We have found six constants, α_j and β_j, which determine the motion of the particle uniquely.

13.3.3 Relations between the canonical constants and orbital elements

Now we shall seek the relations between the canonical constants and orbital elements, since astronomical measurements are usually expressed in terms of these elements. As constants α_1, α_2 and α_3 are contained in expressions L and N, defined by square roots (13.26), we shall start with the examination of the points along the orbit where these square roots are zero.

a) Interpretation of α_1, β_1 *and* α_2. In the first place, specify r_1 as the distance of the perigee, β_1 as the time of passage through the perigee (denote it by τ). At the perigee and apogee we have $\dot{r} = 0$ and, consequently, $p_r = 0$ and $L = 0$ in view of (13.19) and (13.31). Consequently, the square of the function L, given by (13.26), may be expressed as

$$0 = 2\alpha_1 r^2 + 2\mu r - \alpha_2^2 = 2\alpha_1 (r - r_1)(r - r_2) , \tag{13.32}$$

where the roots r_1 and r_2 of this trinomial are the distances of the perigee and apogee, respectively. By putting $v = 0$ and $v = \pi$ in the polar equation of an ellipse (12.9), we obtain $r_1 = a(1-e)$ for the distance of the perigee, and $r_2 = a(1+e)$ for the distance of the apogee, respectively. By inserting these expressions into (13.32), and comparing the coefficients at the individual powers of r, we get

$$\alpha_1 = -\frac{\mu}{2a} , \quad \alpha_2^2 = \mu a(1 - e^2) . \tag{13.33}$$

Constant α_1 is referred to as the constant of the potential energy, since it is equal to the potential energy at distance $r = 2a$ (for mass $m = 1$). It follows that the semi-major axis of the orbit is constant.

Constant α_2 may be expressed, using (12.8) and (12.67), as

$$\alpha_2^2 = \mu a(1 - e^2) = \mu b^2 / a = C^2 , \tag{13.34}$$

C being the double of the areal velocity. Consequently, α_2 is the constant of areas. Note that after multiplying it by m, we obtain the angular momentum of the particle.

b) *Interpretation of* α_3. Now, let us consider the points of the orbit where $\dot{\phi} = 0$, i.e. where the meridional component of the particle velocity is zero. This situation occurs at the points where latitude ϕ attains its maximum or minimum value, i.e. where the absolute value of ϕ is equal to inclination i (see below). If $\dot{\phi} = 0$, then $p_\phi = 0$ according to (13.19), which yields $N = 0$ in view of (13.31). By inserting $N = 0$ and $\cos\phi = \cos i$ into (13.26), we get

$$\alpha_3 = \alpha_2 \cos i . \tag{13.35}$$

Consequently, constant $m\alpha_3$ is the z-component of the angular momentum. It follows from (13.35) that the inclination of the orbit is constant.

Let us derive in a more rigorous way that $\dot{\phi} = 0$ indeed for $\phi = \pm i$ (inclination). Consider relation (12.4), i.e.

$$\sin\phi = \sin i \sin u . \tag{13.36}$$

Using the argument of the latitude, u, as the independent variable instead of time t, and differentiating (13.36) yields

$$\cos\phi \, \mathrm{d}\phi = \sin i \cos u \, \mathrm{d}u . \tag{13.37}$$

It follows that $\mathrm{d}\phi = 0$ for $u = \pm 90°$ (at the "midway" point of the orbit between the nodes). By inserting $u = \pm 90°$ into (13.36) we get $\phi = \pm i$, which further leads to (13.35).

c) *Interpretation of* β_2. In order to simplify the expression for β_2, calculate the last integral in (13.29) by means of substitution (13.36), i.e. replace the integration variable ϕ by u. Thus

$$\int_0^{\phi} \frac{\alpha_2}{N}\,\mathrm{d}\phi = \int_0^{\phi} \frac{\alpha_2}{\sqrt{\alpha_2^2 - \dfrac{\alpha_2^2 \cos^2 i}{\cos^2\phi}}}\,\mathrm{d}\phi = \int_0^{\phi} \frac{\cos\phi\,\mathrm{d}\phi}{\sqrt{1-\sin^2\phi-\cos^2 i}} =$$

$$= \int_0^{\phi} \frac{\cos\phi\,\mathrm{d}\phi}{\sqrt{\sin^2 i - \sin^2 i \sin^2 u}} = \int_0^{\phi} \frac{\cos\phi\,\mathrm{d}\phi}{\sin i \cos u} = \int_0^{u} \mathrm{d}u = u\ ,$$

where we have used (13.26), (13.35) to (13.37). Formula (13.29) then simplifies to read

$$u = \int_{r_1}^{r} \frac{\alpha_2}{rL}\,\mathrm{d}r - \beta_2\ .$$

Specify this expression for the perigee. Here $u = \omega$ (the argument of the perigee), and the integral vanishes as the upper limit is $r = r_1$. We arrive at a very simple interpretation of canonical constant β_2, namely

$$\beta_2 = -\omega\ . \tag{13.38}$$

Hence, the argument of the perigee is constant.

d) *Interpretation of* β_3. The integral in the expression for β_3 in (13.30) can be calculated by means of the substitution

$$\sin\psi = \frac{\tan\phi}{\tan i}\ , \tag{13.39}$$

where ψ is a new variable instead of ϕ. This substitution is analogous to (12.6). By differentiating (13.39), we obtain

$$\cos\psi\,\mathrm{d}\psi = \frac{\mathrm{d}\phi}{\cos^2\phi\ \tan i}\ . \tag{13.40}$$

The integral in (13.30) can now be expressed as

$$\int_0^{\phi} \frac{\alpha_3}{N\cos^2\phi}\,\mathrm{d}\phi = \int_0^{\phi} \frac{\cos i}{\sqrt{1-\dfrac{\cos^2 i}{\cos^2\phi}}}\,\frac{\mathrm{d}\phi}{\cos^2\phi} = \int_0^{\phi} \frac{1}{\sqrt{\dfrac{1}{\cos^2 i}-\dfrac{1}{\cos^2\phi}}}\,\frac{\mathrm{d}\phi}{\cos^2\phi} =$$

$$= \int_0^{\phi} \frac{1}{\sqrt{\tan^2 i-\tan^2\phi}}\,\frac{\mathrm{d}\phi}{\cos^2\phi} = \int_0^{\psi} \frac{\tan i\cos\psi}{\sqrt{\tan^2 i-\tan^2 i\sin^2\psi}}\,\mathrm{d}\psi = \int_0^{\psi}\mathrm{d}\psi = \psi\ , \tag{13.41}$$

where we have used (13.26), (13.35), (13.39) and (13.40). Formula (13.30) now becomes

$$\lambda = \psi - \beta_3\ . \tag{12.42}$$

Specify this formula for the ascending node. Here $\lambda = \Omega$ (the longitude of the ascending node, since λ is measured from the vernal equinox or from another fixed direction in the equatorial plane), and $\phi = 0$, so that also $\psi = 0$ in view of (13.39) or (13.41). Relation (13.42) then again yields a simple interpretation of the last canonical constant, namely

$$\beta_3 = -\Omega\ . \tag{13.43}$$

Thus, the longitude of the ascending node is also constant in the case of the unperturbed motion.

Finally, we can summarise the expressions for the canonical constants in terms of the orbital elements as follows:

$$\begin{aligned} &\alpha_1 = -\frac{\mu}{2a}\ , && \beta_1 = \tau\ , \\ &\alpha_2 = \sqrt{\mu a\left(1-e^2\right)}\ , && \beta_2 = -\omega\ , \\ &\alpha_3 = \sqrt{\mu a\left(1-e^2\right)}\cos i\ , && \beta_3 = -\Omega\ . \end{aligned} \tag{13.44}$$

13.3.4 Solution with a perturbing term

We now assume that the Hamiltonian contains a non-zero perturbing function, R. Then parameters α_j and β_j are not constant, and their time variations are described by Eqs. (13.17), i.e. by

$$\dot{\alpha}_j = -\frac{\partial R}{\partial \beta_j},\quad \dot{\beta}_j = \frac{\partial R}{\partial \alpha_j}\quad \left(j = 1, 2, 3\right)\ . \tag{13.45}$$

These equations solve the problem of perturbed motion in terms of parameters α_j and β_j.

We are, however, interested in the corresponding perturbations of the orbital parameters, since these quantities are usually measured. For this purpose we shall need the expressions for the orbital elements in terms of the canonical constants, i.e. the expressions which are inverse to (13.44). They have the following form:

$$\begin{aligned} a &= -\frac{\mu}{2\alpha_1}, & \tau &= \beta_1, \\ e &= \sqrt{1+2\alpha_1\frac{\alpha_2^2}{\mu^2}}, & \omega &= -\beta_2, \\ i &= \arccos\frac{\alpha_3}{\alpha_2}, & \Omega &= -\beta_3. \end{aligned} \tag{13.46}$$

Firstly, let us calculate the regression of the ascending node, i.e. the derivative $\dot{\Omega}$. Generally,

$$\dot{\Omega} = \sum_{j=1}^{3}\frac{\partial\Omega}{\partial\alpha_j}\dot{\alpha}_j + \sum_{j=1}^{3}\frac{\partial\Omega}{\partial\beta_j}\dot{\beta}_j . \tag{13.47}$$

Since Ω is a function of β_3 only, the previous expression reduces to just one term:

$$\begin{aligned} \dot{\Omega} &= \frac{\partial\Omega}{\partial\beta_3}\dot{\beta}_3 = -\dot{\beta}_3 = -\frac{\partial R}{\partial\alpha_3} = -\frac{\partial R}{\partial i}\frac{\partial i}{\partial\alpha_3} = \\ &= -\frac{\partial R}{\partial i}\frac{-1}{\sqrt{1-\dfrac{\alpha_3^2}{\alpha_2^2}}}\frac{1}{\alpha_2} = \frac{\partial R}{\partial i}\frac{1}{\sqrt{\alpha_2^2-\alpha_3^2}}, \end{aligned} \tag{13.48}$$

where we have used (13.46), (13.45) and again (13.46). Denote by n the mean angular orbital velocity, $n = 2\pi/T$. Kepler's third law (12.68) can then be expressed as

$$\boxed{n^2a^3 = \mu}. \tag{13.49}$$

Using (13.44) and (13.49), formula (13.48) takes the following final form:

$$\boxed{\dot{\Omega} = \frac{1}{na^2\left(1-e^2\right)^{1/2}\sin i}\frac{\partial R}{\partial i}}\,. \tag{13.50}$$

This equation represents one of the *Lagrange equations* for the perturbations of the osculating orbital elements (also referred to as "planetary equations").

In a similar way we can calculate the motion of the perigee in the orbital plane, characterised by velocity $\dot{\omega}$. Analogously to (13.47) and (13.48),

$$\dot{\omega} = \frac{\partial \omega}{\partial \beta_2}\dot{\beta}_2 = -\dot{\beta}_2 = -\frac{\partial R}{\partial \alpha_2} = -\frac{\partial R}{\partial e}\frac{\partial e}{\partial \alpha_2} - \frac{\partial R}{\partial i}\frac{\partial i}{\partial \alpha_2} =$$

$$= -\frac{2\alpha_1\alpha_2}{e\mu^2}\frac{\partial R}{\partial e} - \frac{\alpha_3}{\alpha_2\sqrt{\alpha_2^2-\alpha_3^2}}\frac{\partial R}{\partial i}\,.$$

As opposed to the previous case of $\dot{\Omega}$, here we have obtained two terms, since α_2 appears in (13.46) in two orbital elements, namely in e and i. Again using (13.44) and (13.49), we arrive at the final formula

$$\boxed{\dot{\omega} = \frac{1}{na^2}\left[\frac{\sqrt{1-e^2}}{e}\frac{\partial R}{\partial e} - \frac{\cot i}{\sqrt{1-e^2}}\frac{\partial R}{\partial i}\right]}\,. \tag{13.51}$$

The remaining four Lagrange equations can be derived in a similar way. They can be expressed as

$$\dot{a} = -\frac{2}{n^2 a}\frac{\partial R}{\partial \tau}\,, \tag{13.52}$$

$$\dot{e} = -\frac{1}{na^2 e}\left[\sqrt{1-e^2}\frac{\partial R}{\partial \omega} + \frac{1-e^2}{n}\frac{\partial R}{\partial \tau}\right], \tag{13.53}$$

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{1}{na^2\left(1-e^2\right)^{1/2}\sin i}\left[\cos i\frac{\partial R}{\partial \omega} - \frac{\partial R}{\partial \Omega}\right], \tag{13.54}$$

$$\dot{\tau} = \frac{1}{n^2 a}\left[2\frac{\partial R}{\partial a} + \frac{1-e^2}{ae}\frac{\partial R}{\partial e}\right]. \tag{13.55}$$

Let us also mention some other forms of the Lagrange planetary equations. Firstly, if the mean anomaly v_0 is used as an orbital parameter instead of time τ, Eqs. (13.52), (13.53) and (13.55) must be modified by the substitution

$\mathrm{d}\,v_0 = -n\,\mathrm{d}\,\tau$; see the definition of v_0 in (12.3). Secondly, the form (13.50) to (13.55) of the Lagrange equations cannot be used directly if the osculating plane is in the equatorial plane $(i = 0)$, or if the osculating orbit is circular $(e = 0)$. In these cases, other orbital elements must be used (Burša and Pec, 1993).

The Lagrange equations generally solve the problem of perturbed motion. However, to be able to apply them, we must know the explicit form of perturbing function R and, moreover, express this function in terms of the orbital elements, with respect to which the differentiation is performed. A very important special form of the perturbing function will be considered in the next section.

13.4 Perturbations of Elliptical Orbits Due to the First Term of the Perturbing Potential

Consider again the series for the external gravitational potential of the Earth in the form (12.24), i.e.

$$V = -\frac{GM}{r}\left[1 - J_2\left(\frac{a_0}{r}\right)^2 P_2\left(\sin\phi\right) + \ldots\right], \tag{13.56}$$

where a_0 is the equatorial radius of the Earth, J_2 is a zonal geopotential parameter, and $P_2\left(\sin\phi\right) = \frac{3}{2}\sin^2\phi - \frac{1}{2}$.

In addition to the main potential term, $-GM/r$, let us consider only the first term of the perturbing potential, containing zonal parameter J_2. The perturbing Hamiltonian then reads (we have put $m = 1$)

$$H_1 = \frac{\mu J_2 a_0^2}{2r^3}\left(3\sin^2\phi - 1\right), \tag{13.57}$$

where we have again put $\mu = GM$. To replace spherical coordinates r, ϕ by orbital parameters, relations (12.4) and (12.9) may be used, i.e.

$$\sin\phi = \sin i \sin u\ , \tag{13.58}$$

$$r = \frac{a\left(1 - e^2\right)}{1 + e\cos v}\ , \tag{13.59}$$

where $v = u - \omega$. Putting $\mu = n^2 a^3$ in view of (13.49), perturbing function $R = -H_1$ may be expressed as

$$R = -\frac{n^2 a^3 J_2 a_0^2}{2r^3}\left(3\sin^2 i \sin^2 u - 1\right), \tag{13.60}$$

where r is given by (13.59).

Since the first term of the perturbing potential mainly causes a regression of the nodal line of the orbital plane and a rotation of the perigee in this plane, here we shall deal only with the perturbations of elements Ω and ω.

By inserting the perturbing function (13.60) into formula (13.50) for $\dot{\Omega}$, we get

$$\dot{\Omega} = -3na\frac{J_2 a_0^2 \cos i}{\sqrt{1-e^2}}\frac{\sin^2 u}{r^3}. \tag{13.61}$$

This formula yields the instantaneous velocity $\dot{\Omega}$. This velocity depends on the position of the satellite, as the last fraction varies with time.

We are now seeking the mean value of $\dot{\Omega}$, averaged over one revolution. If the osculating orbit is circular $(e = 0)$, the averaging is easy to perform. In this case, $\overline{\sin^2 u} = \frac{1}{2}$ and $r = a$, so that formula (13.61) simplifies to formula (12.39), which was derived using another method earlier; only different notations are used for the angular velocity, $n = \omega$.

In the case of an elliptical orbit, accurate averaging is more complicated, since motion along the ellipse is not uniform as a consequence of Kepler's second law. We must transform angles u and v into quantities which vary with time uniformly. It can then be proved that the mean value of $\left(\sin^2 u\right)/r^3$ is $\frac{1}{2}a^{-3}\left(1-e^2\right)^{-3/2}$; we shall not perform this complicated calculation here. For the mean value of $\dot{\Omega}$ we then obtain

$$\boxed{\dot{\Omega}_{mean} = -\frac{3}{2}nJ_2\frac{a_0^2}{a^2}\frac{\cos i}{\left(1-e^2\right)^2}}. \tag{13.62}$$

The perturbation of the argument of the perigee, $\dot{\omega}$, can be expressed by means of (13.51), (13.60) and (13.59). After averaging, one gets

$$\boxed{\dot{\omega}_{mean} = \frac{3}{2}nJ_2\frac{a_0^2}{a^2\left(1-e^2\right)^2}\left(2-\frac{5}{2}\sin^2 i\right)}. \tag{13.63}$$

This velocity is zero if the expression in the last parentheses is zero, which occurs for inclination $i = 63.4°$. It follows from (13.63) that if

$$0 \le i < 63.4° \Rightarrow \dot{\omega} > 0 \ ,$$

$$63.4° < i \le 90° \Rightarrow \dot{\omega} < 0 \ .$$

(Note that especially the first Soviet satellites had inclinations close to about $65°$, in order to fly over the territory of the former Soviet Union).

If the osculating elements and perturbations $\dot{\Omega}_{mean}$ or $\dot{\omega}_{mean}$ of an orbit are measured, formulae (13.62) and (13.63) then enable us to determine the geopotential parameter J_2.

Since, the geocentric gravitational constant $\mu = GM$ is now known with a high accuracy, we can express the mean angular velocity n from Kepler's third law (13.49), i.e. $n = \sqrt{\mu/a^3}$. Formulae (13.62) and (13.63) then take the form

$$\dot{\Omega}_{mean} = -\frac{3}{2}\sqrt{\mu}\, J_2 \frac{a_0^2}{a^{7/2}} \frac{\cos i}{\left(1-e^2\right)^2} \ ,$$

$$\dot{\omega}_{mean} = \frac{3}{2}\sqrt{\mu}\, J_2 \frac{a_0^2}{a^{7/2}\left(1-e^2\right)^2}\left(2 - \frac{5}{2}\sin^2 i\right) \ . \tag{13.64}$$

In these formulae, let us consider the same numerical values of μ, a_0 and J_2 as in Tabs. 12.1 and 12.2. For an equatorial orbit, closely above the Earth's surface, the perturbations reach values of nearly

$$\left|\dot{\Omega}\right|_{max} = 10°/\text{day} \ , \quad \dot{\omega}_{max} = 20°/\text{day} \ . \tag{13.65}$$

These large changes of orbital elements can easily be measured astronomically, especially if a cumulative effect, e.g. after many months, is determined. It demonstrates that observations of the motions of the node and perigee represent powerful methods of determining geopotential parameter J_2 and, consequently, of determining the flattening of the Earth.

Let us remind the reader of the three principal methods of determining the *flattening* of the Earth, namely (see Section 7.8):

1) the geometric method based on *arc measurements*;
2) the gravimetric method based on *Clairaut's theorem*;
3) *satellite methods*.

The analyses in this chapter have clearly demonstrated the advantages of satellite methods for these purposes.

13.5 Perturbations of Orbits Due to Other Terms of the Perturbing Potential

13.5.1 Effects of even zonal terms

From an algebraic point of view, coefficient J_2 contributes to the motion of the node and perigee, because the mean value of $\sin^2 u$ in $P_2(\sin\phi)$ is not zero (but about $1/2$). Consequently, also the other even terms will contribute to this motion. A more general theory must take these terms into account, e.g., in the following form:

$$\dot{\Omega}_{mean} = a_2 J_2 + a_4 J_4 + a_6 J_6 + \dots ,$$

$$\dot{\omega}_{mean} = b_2 J_2 + b_4 J_4 + b_6 J_6 + \dots .$$

Expressions a_i and b_i can be derived from the theory. Since they depend in a different way on the elements of the orbit, especially on inclination i, observations of many orbits with different orbital elements can be used to determine the individual coefficients J_2, J_4, J_6, etc. The numerical values of these coefficients can be found in Tab. 13.1 below.

13.5.2 Effects of odd zonal terms

The odd zonal terms of the potential mainly cause *long-period oscillations of the eccentricity and inclination*. For example, the distance of the perigee oscillates with amplitudes which may reach several kilometres. The differences between the distance of the perigee at the equator $(\omega = 0)$ and at the apex $(\omega = \pm 90°)$ are usually measured. If the amplitudes of this effect are measured for several satellites (especially with different inclinations), coefficients $J_3, J_5, J_7, \dots$, can be determined.

The Legendre polynomials of odd degrees, i.e. $P_3(\sin\phi)$, $P_5(\sin\phi)$, etc., are odd functions of geocentric latitude ϕ. This means that the odd coefficients $J_3, J_5, \dots$, characterise the asymmetries of the Earth's gravitational field with respect to the equatorial plane. Such asymmetries were not known and were not considered in the pre-satellite era. For example, the harmonic terms for $n = 2$ and $n = 4$ were usually included in the normal gravity field, but the term for $n = 3$ not. Consequently, when satellite observations confirmed the non-zero value of coefficient J_3, the problem of the Earth's figure was widely discussed even outside technical circles. Since the equipotential surfaces corresponding to this term resemble of a pear in shape, the concept of a "pear-shaped Earth" became very popular. However, later observations revealed that also further coefficients in the geopotential have values comparable with J_3. Hence, it

became evident that the geoid has a very irregular shape, where the small pear-shaped component does not predominate over other components. The notion of a pear-shaped Earth was then abandoned.

In the end, we should specify the periods of the oscillations mentioned above. Since a usual shift of the perigee is about 5° per day (the value given in (13.65) is an extreme), the period comes out at about 2 months.

13.5.3 Effects of tesseral terms

The tesseral terms, which depend on geographical longitude λ, cause short-period oscillations of various orbital elements (with periods of one day, half a day, etc.). Determining the corresponding coefficients requires accurate measurements, e.g., observations of the Doppler effect or laser measurements. Namely, the measurements of tesseral coefficients are complicated by the Earth's rotation, which leads to averaging the effects dependent on the geographical longitude. Consequently, instantaneous parameters of orbits must be measured for these purposes.

Another option is to study various resonance phenomena. We say that a satellite is in resonance with the Earth's rotation if, after several revolutions, it again orbits over the same places. In this case, the effects from some places are accumulated, which facilitates the determination of the corresponding tesseral coefficients.

13.6 Models of the Earth's Gravitational Field

Models of the Earth's gravitational potential (geopotential models) have been rapidly developed since the 1960's. Here we shall give only an overview of these models.

13.6.1 Geopotential models evolution

The evolution of geopotential models was briefly described by Blitzkow and Mello (1997) as follows:

"An import and decisive contribution to geopotential models came in 1966 with the publication of Smithsonian Astrophysical Observatory (SAO) model complete to degree and order 8. The model really fixed the zonal coefficients obtained previously and derived the others. New models followed in 1973, 1974 and 1977 with the denomination of SE (Standard Earth).

Goddard Space Flight Center (GSFG) started a parallel effort from 1976 with the denomination GEM (Goddard Earth Model). The first models of the series up to GEM9 used again only satellite observations. From GEM10 mean gravity anomalies have been included as well as altimeter data in a sequential improvement with models GEM10, GEM10A, GEM10B and GEM10C.

The Groupe de Recherches de Géodésie Spatiale and Sonderforschungsbereich carried out a different effort with the publication of the GRIM model series."

In 1994 the National Imagery and Mapping Agency (NIMA) and NASA/Goddard Space Flight Center signed a Memorandum of Understanding for Joint Gravity Field and Geoid Improvement Project. The primary goal of the project was to set up a new geopotential model and its associated global geoid. Some characteristics of this model, referred to as the EGM96 model, are given below.

13.6.2 An example of geopotential models

Consider the gravitational potential of the Earth in the form

$$V = -\frac{GM}{r}\left[1+\sum_{n=2}^{\infty}\left(\frac{a_0}{r}\right)^2\sum_{m=0}^{n}\left(J_n^m\cos m\lambda + S_n^m\sin m\lambda\right)P_n^m\left(\sin\phi\right)\right], \quad (13.66)$$

where r is the radial distance from the Earth's centre, ϕ is the geocentric latitude, λ the longitude (positive to the east).

Table 13.1: GEM-T2 geopotential (Stokes') parameters: conventional J_n^0; fully normalised $\bar{J}_n^m$ and $\bar{S}_n^m$.

n	$J_n^0\left(10^{-6}\right)$	m	$\bar{J}_n^m\left(10^{-6}\right)$	$\bar{S}_n^m\left(10^{-6}\right)$
2	-1 082.6265	0	-484.1653	
		1	-0.0017	0.0012
		2	2.4390	-1.4001
3	2.5321	0	0.9570	
		1	2.0308	0.2496
		2	0.9035	-0.6190
		3	0.7215	1.4137
4	1.6197	0	0.5399	
		1	-0.5353	-0.4741
		2	0.3483	0.6640
		3	0.9913	-0.2014
		4	-0.1894	0.3090
5	0.2278	0	0.0687	
		1	-0.0608	-0.0950
		2	0.6561	-0.3241
		3	-0.4519	-0.2171
		4	-0.2950	0.0514
		5	0.1719	-0.6691

Table 13.1 gives the numerical values of the geopotential coefficients for model GEM-T2. This model, containing coefficients up to degree and order $n = m = 36$, was derived from laser tracking of 11 satellites, especially geodetic satellites (Marsh et al., 1990). Table 13.1 contains only the coefficients for low degrees $n \leq 5$; for a more complete set of these coefficients we refer the reader, e.g., to Burša and Pec (1993). It should be noted that opposite signs have been used with the zonal terms in (13.56) and (13.66). Consequently, coefficient J_2 in (13.56) is positive, whereas the same coefficient in (13.66), denoted by J_2^0, is negative.

The last two columns of Tab. 13.1 contain the so-called fully normalised coefficients which are defined by the relations (Burša and Pec, 1993)

$$\begin{Bmatrix} J_n^m \\ S_n^m \end{Bmatrix} = \sqrt{\frac{(2-\delta_{0m})(2n+1)(n-m)!}{(n+m)!}} \begin{Bmatrix} \bar{J}_n^m \\ \bar{S}_n^m \end{Bmatrix}, \tag{13.67}$$

where J_n^m, S_n^m are conventional parameters, and δ_{0m} is Kronecker's symbol.

13.6.3 EGM96 geopotential model

The latest, very detailed Earth's geopotential model was presented officially in 1996 as the EGM96 model (Rapp and Nerem, 1996). The model incorporated existing and new satellite data, and also surface gravity data.

The satellite tracking data from more than 30 different satellites, spanning many years in time, served as basic data for the EGM96 model. Moreover, the data from the new satellites LAGEOS 1, ERS-1, Starlette and several others were also included. The surface gravity data were based on updated mean anomalies in the grid 30×30 minutes.

The EGM96 model is complete to degree and order $n = m = 360$. For details on this model we refer the reader, e.g., to Blitzkow and Mello (1997).

Chapter 14

The Figure of the Real Earth's Surface

Potential fields in physics are frequently described by means of force lines, or equipotential surfaces. Since the geoid is one of the equipotential surfaces of the Earth's gravity field, the study of its figure represents an important problem of gravimetry. The undulations of the geoid are associated with irregularities in the density distribution within the Earth, and these irregularities are associated with some processes within the Earth. Consequently, by studying the undulations of the geoid we may study the corresponding physical parameters and processes, such as the stress state within the Earth, or convective currents. Moreover, in geophysics, the figure of the Earth is understood to be the figure of the geoid. This further emphasises the importance of this surface. In other words, if the figure of the geoid were known, the problem of the figure of the Earth would be solved completely from the geophysical point of view.

On the contrary, the geoid has no principal importance for geodesy, whose final task is to determine the figure of the real Earth's surface. To accomplish this aim, it is sufficient to determine the heights of the real surface above the geoid, i.e. the heights above sea level, providing the figure of the geoid is known. We shall deal with these problems in Section 14.1. Nevertheless, if we were able to determine the figure of a surface other than the geoid, and the heights above it, this new surface would suit equally well as the geoid. Molodenskii et al. (1960) proposed to use the "quasigeoid" instead of the geoid, because the quasigeoid can be determined with greater accuracy (Section 14.3).

14.1 Heights above Sea Level

The height of a point above sea level may be defined as the distance along the plumb line between this point and the geoid. This height is referred to as the *orthometric height*. The orthometric height of point B in Fig. 14.1 is its distance from the corresponding point C on the geoid. Note that it is not necessary to distinguish between the distance along the plumb line and the straight-line distance of points B and C, because the curvature of plumb lines is usually small.

In practice, the height of point B above sea level is determined by levelling from an initial point at sea level, O, along the Earth's surface; see levelling traverse OAB in Fig. 14.1. In the course of levelling, the levelling instrument is adjusted to the horizontal plane, i.e. to the plane which is tangent to the local equipotential surface of the gravity field. However, these equipotential surfaces are not parallel in the geometrical sense. They converge due to the Earth's

flattening, and due to the anomalous distribution of masses. This convergence causes errors in accurate levellings.

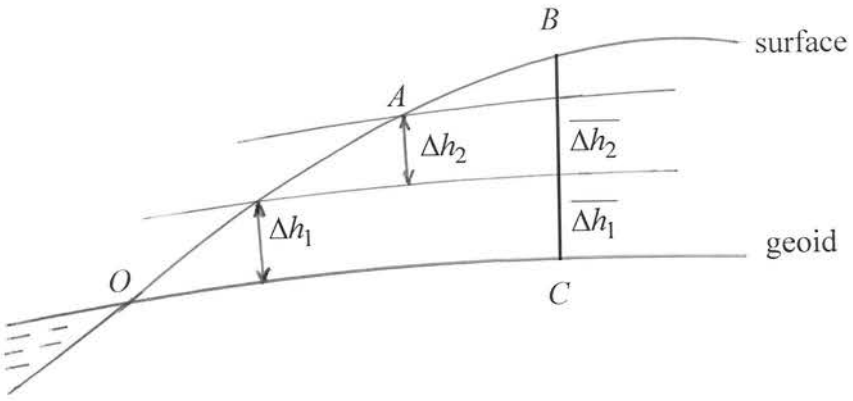

Figure 14.1: Determination of the height above sea level.

Denote Δh_i the superelevation measured in the i-th levelling step, and $\overline{\Delta h_i}$ the corresponding increment along plumb line CB. Generally $\Delta h_i \neq \overline{\Delta h_i}$ as the consequence of the convergence of the equipotential surfaces. The orthometric height of point B is $h_o(B) = \sum \overline{\Delta h_i}$, and its height determined by levelling is $h_l(B) = \sum \Delta h_i$. However, the latter value depends on the path of levelling, and generally

$$\sum_i \Delta h_i \neq \sum_i \overline{\Delta h_i} \ . \tag{14.1}$$

Hence, height h_l differs slightly from orthometric height h_o. Moreover, height h_l is not defined uniquely.

To obtain unique values, it is necessary to express the heights in terms of the gravity potential. Since the corresponding increments in the gravity potential are equal, i.e.

$$g_i \Delta h_i = \overline{g}_i \overline{\Delta h_i} \ , \tag{14.2}$$

integral $\int_O^B g \, \mathrm{d}h$ is already independent of the path of integration. This integral can be expressed as

$$\int_O^B g \, \mathrm{d}h = \int_C^B \overline{g} \, \overline{\mathrm{d}h} = g_m \int_C^B \overline{\mathrm{d}h} = g_m h_o(B) \ , \tag{14.3}$$

where g_m is the mean value of the gravity acceleration along CB. In this way we arrive at the following formula for the *orthometric height*:

$$\boxed{h_o(B) = \frac{1}{g_m}\int_O^B g\,\mathrm{d}h}\,. \tag{14.4}$$

The integral in the latter formula can be calculated if the levelling is accompanied by gravity measurements along the levelling traverse. However, in general, the mean gravity along the plumb line, g_m, is not known. Hence, it is not possible to determine the orthometric height accurately.

It is, therefore, necessary to introduce other heights which can be determined more accurately. Replace g_m in (14.4) by the mean value of the normal gravity, γ_m, along line CB. Value γ_m can be determined, since the normal gravity can be calculated for any height; see formula (7.57). In this way, instead of orthometric height h_0 we arrive at

$$\boxed{h_N(B) = \frac{1}{\gamma_m}\int_O^B g\,\mathrm{d}h}\,. \tag{14.5}$$

This height is referred to as the *normal height*.

In plains, the normal height may differ from the orthometric height only by several centimetres, but in the mountains the difference may exceed one metre.

If orthometric heights are plotted from the Earth's surface downward, the geoid is obtained. If the same procedure is applied to normal heights, one obtains a different surface, called the *quasigeoid*.

The quasigeoid represents an approximation of the geoid. Both the surfaces coincide on the oceans, but they may differ by several centimetres in lowlands and by about one metre in high mountains, which corresponds to the differences between the orthometric and normal heights. Note that the quasigeoid is not an equipotential surface.

By replacing the mean value of the normal gravity by a constant (usually by the normal gravity on the reference ellipsoid at latitude 45°, i.e. for $h = 0$ and $\phi = 45°$), we arrive at another approximation, called the *dynamic height*:

$$\boxed{h_D(B) = \frac{1}{\gamma_{45°}}\int_O^B g\,\mathrm{d}h}\,. \tag{14.6}$$

Dynamic heights may differ from normal heights by several metres. Dynamic heights are frequently used in practical applications, e.g., in civil engineering.

14.2 Problems in the Classical Method of Determining the Earth's Surface

The shape of the real Earth's surface is very complicated and, as opposed to the spheroid or reference ellipsoid, cannot be described by a few parameters. Consequently, the usual procedure of determining the real Earth's surface consists in the following three steps:

1) The determination of a suitable reference ellipsoid, i.e. the determination of its semi-major axis a, flattening α, and its orientation in the Earth (described by the deflections of the vertical, and by an azimuth at an initial point).
2) The determination of the distances between the ellipsoid and the geoid, e.g., by means of Stokes' theorem.
3) The determination of the heights of the real surface above the geoid (above sea level).

In this process, the geoid plays the role of an auxiliary surface which only facilitates the transition from the ellipsoid to the real Earth's surface. The geoid only divides the height above the ellipsoid (ellipsoidal height) into two parts, namely the undulation of the geoid, ζ_o, and the height above sea level, h_o (Fig. 14.2). The height above sea level is determined by levelling, and the undulation by solving a boundary problem of potential theory, formulated on the geoid.

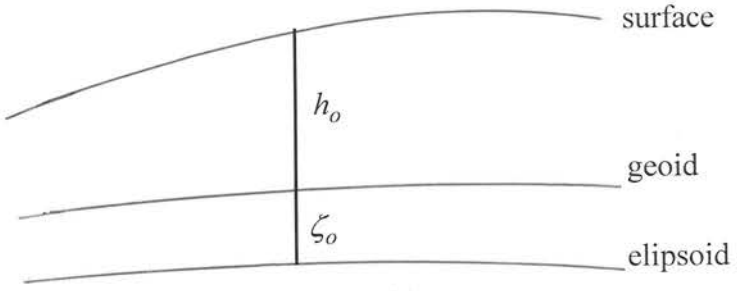

Figure 14.2: Division of the ellipsoidal height into the undulation of the geoid, ζ_o, and orthometric height h_o.

There are several principal reasons which prevent steps 2 and 3, mentioned above, to be performed with great accuracy. Namely, there are certain limitations to Stokes' theorem in determining the geoid, and problems with reducing surface gravity data to the geoid, as we have already seen in Chapters 8 and 10. Consequently, the undulations of the geoid cannot be determined accurately. To determine them accurately, it would be necessary to know the density distribution and actual gravity acceleration between the Earth's surface and the geoid. However, these quantities are not generally known. Molodenskii

(1945) even proved that it is impossible to determine the geoid by any method only from measurements on the real surface of the Earth.

Moreover, in this chapter we have also encountered problems in determining heights above sea level. Therefore, we have arrived at the conclusion that neither the undulations of the geoid, nor heights above sea level can be determined accurately.

14.3 Principles of Molodenskii's Method for Determining the Figure of the Earth's Surface

Molodenskii was the first to recognise the auxiliary role of the geoid in the process of determining the real Earth's surface. To solve the above-mentioned problems, he proposed the following fundamental changes:

1) to use the *quasigeoid* instead of the geoid;
2) to determine the undulations of the quasigeoid by integrating over another surface which is *close to the Earth's surface*.

We shall discuss the individual modifications only briefly. For details we refer the reader to Molodenskii et al. (1960) and Pick et al. (1973).

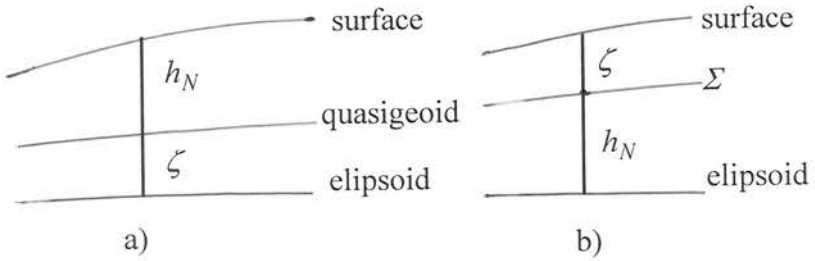

Figure. 14.3: Molodenskii's modifications: a) the division of the ellipsoidal height by the quasigeoid into the undulation of the quasigeoid, ζ, and normal height h_N; b) the construction of surface Σ by plotting the normal height from the ellipsoid.

In the usual approach (Fig. 14.2), neither the undulations of the geoid, nor the orthometric height can be determined accurately. If the geoid is replaced by the quasigeoid (Fig. 14.3a), at least normal heights can be determined accurately.

Nevertheless, the new division of the ellipsoidal height into the undulation of the quasigeoid (distance of the quasigeoid from the ellipsoid, ζ) and normal height h_N will be only formal until a new method for calculating undulations ζ is also developed. This problem was also solved by Molodenskii et al. (1960). Their method is based on the following considerations.

The undulations of the geoid, and also of the quasigeoid, are of the order of tens of metres. Only exceptionally do they amount to 100 m. However, heights above sea level amount to hundreds and thousands of metres. Therefore, the

geoid and quasigeoid in continental areas are generally closer to the ellipsoid than to the Earth's surface. Consequently, the reductions of the gravity measurements from the Earth's surface to the geoid (required in Stokes' theorem and other classical approaches) represent extrapolations to rather large depths. This may lead to considerable errors. Formulation of the boundary problem on the quasigeoid would encounter similar problems.

For this reason, Molodenskii et al. (1960) proposed to use another surface, Σ, which is closer to the Earth's surface (Fig. 14.3b). This surface is obtained by reverting the heights plotted from the ellipsoid toward the Earth's surface. If distance ζ is plotted above the ellipsoid, the quasigeoid is obtained (Fig.14.3a). However, if normal height h_N is plotted, we obtain surface Σ. Note that surface Σ is sometimes referred to as the terroid, hypsometric surface, or the surface of the first approximation. Molodenskii et al. (1960) succeeded in formulating the corresponding boundary problem on surface Σ, i.e. in calculating undulations ζ of the quasigeoid by integrating over surface Σ. The disturbing potential in their theory is sought in the form of the potential of a single layer which is spread on surface Σ. To obtain the unknown density of this layer, an integral equation must be solved. Its solution is usually carried out by the method of a small parameter (Pick et al., 1973).

Since surface Σ is close to the Earth's surface, only small gravity reductions are needed. Consequently, much greater accuracy can be attained in comparison with the classical approaches. In practice, the integration over surface Σ is even replaced by integration over the real Earth's surface, as a first approximation. In this case, no gravity reductions are introduced. The results obtained in the first approximation are usually so accurate that no further approximations are needed.

In spite of the great progress achieved by Molodenskii's theory, some theoretical and practical problems still remain. For example, the integrals in Molodenskii's theory belong to a category of improper integrals which are difficult to evaluate. Consequently, various modifications and combinations of methods have also been developed.

Boundary problems for partial differential equations belong to the traditional problems in textbooks of mathematical physics. The reader may, therefore, ask the questions: Why was the boundary problem for the external gravity field not already solved a long time ago, and why is it so difficult even now? The answer is that our problem of the gravity field is much more complicated than the classical problems. In particular, in the classical boundary problems it is assumed that the figure of the boundary is known. On the contrary, in our problem we must calculate not only the external gravity field, but simultaneously also the unknown shape of the boundary (of the Earth's surface). Our problem thus belongs to the category of problems which are sometimes referred to as boundary problems with a free boundary. Let us mention another typical problem from this category, namely heat transport in an iceberg which floats in water; as the iceberg melts, the interface between the

ice and water gradually changes. Universal mathematical methods for solving these difficult problems have not been developed yet.

14.4 Satellite Altimetry

The introduction of satellites has added immensely to our knowledge of the external gravitational and electromagnetic fields of the Earth, motions of the Earth, and many other parameters and processes which are studied in astronomy, meteorology and geodesy.

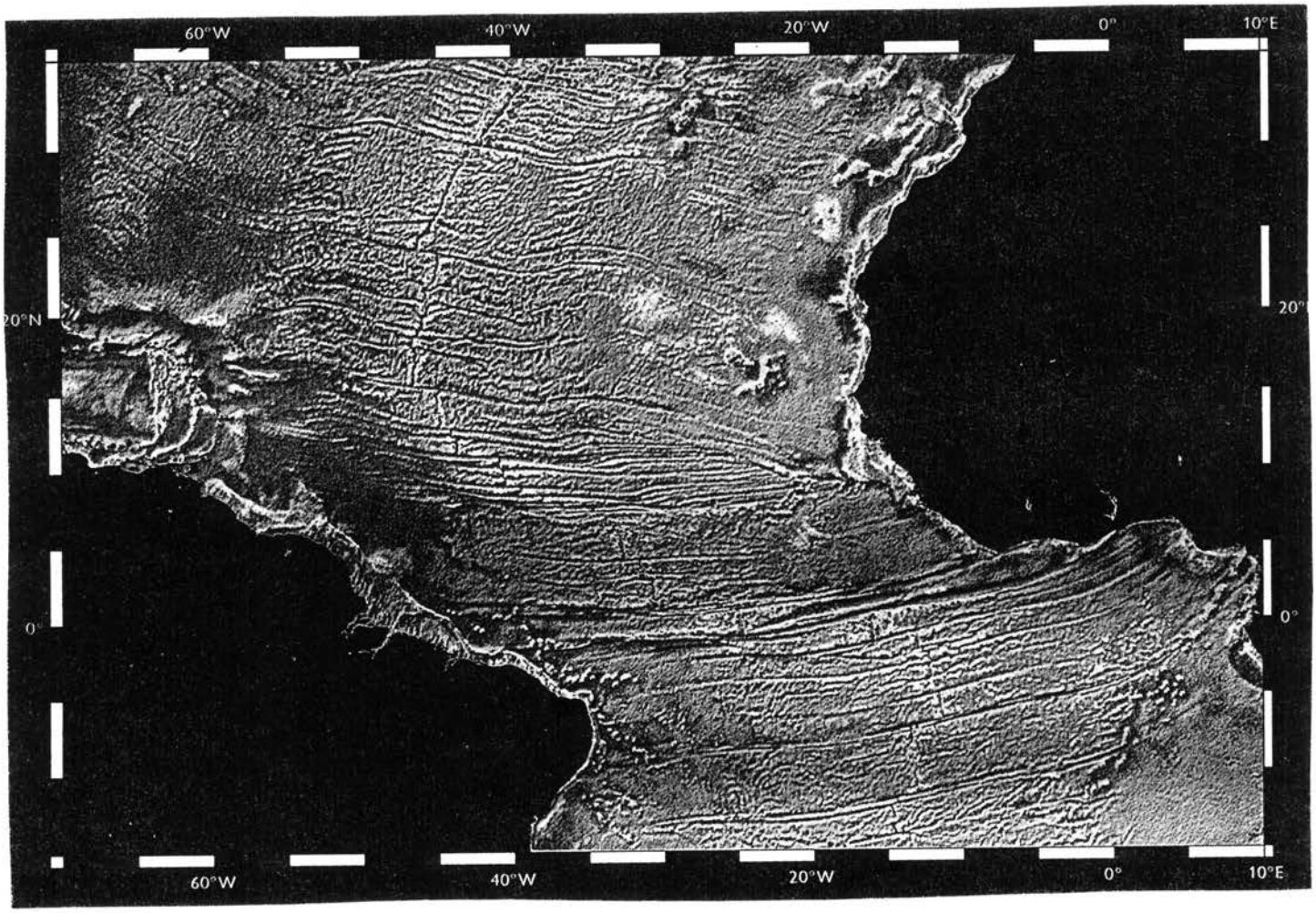

Figure 14.4: Shape of the sea level between Africa and South America as determined by satellite altimetry. (After Kunzig (1996)).

Among the direct geodetic applications of satellites, we have already mentioned geometric methods, in which the satellite, used as a passive object, is simultaneously observed from several places (using optical methods, laser telemeters, etc.). Another important technique of space geodesy is *satellite*

altimetry. In this method, a radar on board a satellite, the orbit of which is known with great accuracy, measures the distances between the satellite and the Earth's surface.

Satellite altimetry has led to dramatic advances in recent years in investigating the shape of the surface of the oceans and seas (the water level is a good reflector of electromagnetic waves). Since the electromagnetic impulse is reflected from an area of several square kilometres, small-scale irregularities of the surface, such as water waves, are smoothed in these measurements. Figure 14.4 shows the results of satellite altimetry in the region between Africa and South America. It can be seen in the figure that the shape of the sea level is closely related to the structural features of the Earth's crust in the region. In particular, the fracture zones that cut across the Midatlantic Ridge show the path of the separating continents. It is evident that such altimetric data will be very valuable in future studies of the structure and dynamics of the Earth's crust, and the dynamics of the ocean.

Satellite altimetry cannot be applied on continents, because the reflected signals are too weak (due to lower reflectivity and larger scattering of the continental surface). Therefore, other satellite techniques must be sought for these purposes. One possibility is mentioned below.

14.5 Global Positioning System (GPS)

The *Global Positioning System*, abbreviated GPS, is a satellite system which is designed to determine accurately position on the Earth's and in its neighbourhood. The system, completed in the first half of the 1990's, consists of three parts:

1) system of satellites (see below);
2) system of ground stations, which observe the satellites, determine the parameters of their orbits, and transmit these data to the satellites;
3) receivers of individual users, which receive the satellite signals, and calculate the coordinates of the points of observation.

The basic technical data of the satellite system are as follows:

- 24 satellites;
- 6 orbital planes (4 satellites in each plane);
- inclination of the orbits of 55° ;
- satellite height of 20 200 km, orbital period of 12 hours;
- diameter of the satellite of about 2.5 m;
- two frequency bands:
  $f_1 = 1.5$ GHz, $\lambda_1 = 19$ cm ,
  $f_2 = 1.2$ GHz, $\lambda_2 = 24$ cm ;
- coding and phase modulation;
- maximum accuracy in determining position:
  horizontal distances ~ 1 cm,
  vertical distances ~ 2–3 cm.

The satellite signals can be analysed in several regimes. A simple, fast processing of satellite signals yields an accuracy of several tens of metres. Such an accuracy in determining the positions is sufficient for the navigation of airplanes, ships, military vehicles, and for similar purposes. If the signals are processed in more detailed, an accuracy of several centimetres can be attained. These accurate measurements may find applications in solving many geodetic problems. Moreover, if these accurate measurements are repeated after a time, slow motions and deformations of the Earth's crust can be studied. In this way, we have now obtained quite new technical tools for direct measurements of slow geological motions, such as the continental drift, recent motions, deformations preceding earthquakes or volcanic eruptions, and some others.

The GPS, although commissioned only several years ago, now finds wide practical and scientific applications. We may expect its importance to increase in the forthcoming years and decades.

References

Arfken, G., 1970. Mathematical Methods for Physicists. (Second Edition). Academic Press, New York.

Båth, M., 1979. Introduction to Seismology. (Second, Revised Edition). Birkhäuser Verlag, Basel.

Blitzkow, D., and Mello, S., 1997. EGM96 geopotential model and geophysics. In: 5th International Congress of the Brazilian Geophysical Society, Vol. I, 45-46. São Paulo.

Buchar, E., 1958. Motion of the nodal line of the second Russian Earth satellite (1957β) and flattening of the Earth. Nature **182**, 198-199.

Burša, M., et al., 1995. Report of I.A.G. Special Commission SC3 "Fundamental Constants". XXIst I.A.G. General Assembly, Boulder.

Burša, M., and Pec, K., 1993. Gravity Field and Dynamics of the Earth. Springer-Verlag, Berlin.

Clairaut, A. C., 1743. Théorie de la figure de la terre. Paris.

Doodson, A. T., 1921. The harmonic development of the tide-generating potential. Proc. Roy. Soc. A 100.

Fischer, I., 1975. The figure of the Earth - changes in concepts. Geophysical Surveys **2**, 3-54.

Garland, G. D., 1965. The Earth's Shape and Gravity. Pergamon Press, New York.

Griffin, W. R., 1949. Residual gravity in theory and practice. Geophysics, **14**, 39-56.

Grushinskii, N. P., 1965. Teoriya figury Zemli. (1-e izd.). Nauka, Moskva. (Theory of the Figure of the Earth; 1st Ed.; in Russian).

Grushinskii, N. P., 1976. Teoriya figury Zemli. (2-e izd.). Nauka, Moskva. (Theory of the Figure of the Earth; 2nd Ed.; in Russian).

Grushinskii, N. P., 1983. Osnovy gravimetrii. Nauka, Moskva. (Foundations of Gravimetry; in Russian).

Hayford, J. F., 1909. The Figure of the Earth and Isostasy, from Measurements in the United States. U.S. Coast and Geodetic Survey, Washington.

Heiskanen, W. A., and Vening Meinesz, F. A., 1958. The Earth and Its Gravity Field. McGraw-Hill, New York.

Helmert, F. R., 1884. Die mathematischen und physikalischen Theorien der höheren Geodäsie. B. G. Teubner, Leipzig. (Band I, Leipzig 1880; Band II, Leipzig 1884).

Ibrmajer, J., 1978. Tíhové mapy CSSR a jejich geologická interpretace. Geophysics N. E., Brno. (Gravity Maps of Czechoslovakia and Their Geological Interpretation; theses, in Czech).

Kalvoda, J., 1981. Vývoj vesmíru ve vztahu k historii Zeme. Pokroky matematiky, fyziky a astronomie, XXVI, 181-192. (Evolution of the Universe in the relation to the history of the Earth; in Czech).

Kittel, Ch., Knight, W. D., and Ruderman, M. A., 1962. Mechanics. (Berkeley Physics Course, Vol. 1). McGraw-Hill, New York.

Koudelková, G., 1993. Doba ledová. VTM **3/93**, 41-43. (Ice Age; in Czech).

Kunzig, R., 1996. The sea floor from space. Discover, March, 60-64.

Love, A. E. H., 1909. The yielding of the Earth to disturbing forces. Proc. R. Soc. London, A82, 73-88.

Mareš, S. a kol., 1979. Uvod do uzite geofyziky. SNTL, Praha. (Translation into English: Mareš, S., et al., 1984. Introduction to Applied Geophysics. D. Reidel Publ. Comp., Dordrecht).

Marsch, J. G., et al., 1990. The GEM-T2 gravitational model. J. Geophys. Res. **95**, 22043-22071.

Matveev, A. N., 1986. Mekhanika i teoriya otnositelnosti. (2-e izd.). Vysshaya shkola, Moskva. (Mechanics and the Theory of Relativity; 2nd Ed., in Russian).

Melchior, P., 1966. The Earth Tides. Pergamon Press, Oxford.

Melchior, P., 1983. The Tides of the Planet Earth. Pergamon Press, Oxford.

Molodenskii, M. S., 1945. Osnovnye voprosy geodezicheskoi gravimetrii. Tr. CNIIGAiK, vyp. **42**, Geodezizdat, Moskva. (Fundamental Problems of Geodetic Gravimetry; in Russian).

Molodenskii, M. S., Eremeev, V. F., and Yurkina, M. I., 1960a. Metody izucheniya vneshnego gravitatsionnogo polya i figury Zemli. Tr. CNIIGAiK, vyp. **131**, Moskva. (Methods of Studying the Outer Gravity Field and Figure of the Earth; in Russian).

Molodensky, M. S., Eremeev, V. F., and Yurkina, M. I., 1960b. Methods for Study of the External Gravitational Field and Figure of the Earth. Moscow. (Translation published by Israel Program for Scientific Translations, Jerusalem, 1962).

Moritz, H., 1979. Report of Special Study Group No. **5.39** of IAG "Fundamental Geodetic Constants". XVIIth General Assembly of IUGG, Canberra 1979.

Novotný, O., 1982. On the addition theorem for Legendre polynomials. In: Travaux Geophys., **30**, 33-45. Academia, Prague.

Parasnis, D. S., 1997. Principles of Applied Geophysics. (Fifth Edition). Chapman and Hall, London.

Pick, M., 1984. The gravitational effect of bodies with variable density. Studia geoph. et geod., **28**, 381-392.

Pick, M., 1987. Vybrané kapitoly z gravimetrie. PrF UK, Praha. (Selected Chapters from Gravimetry; lecture notes, in Czech).

Pick, M., Pícha, J., and Vyskocil, V., 1973. Theory of the Earth's Gravity Field. Academia, Prague, and Elsevier Scientific Publishing Company, Amsterdam.

Rapp, R. H., and Nerem, R. S., 1996. A joint GSFC/DMA project for improving the model of the Earth's gravitational field. GraGeoMar96 Symposium, Tokyo.

Rasmussen, R., and Pedersen, L. B., 1979. End correction in potential field modeling. Geophys. Prosp., **27**, 749-760.

Resal, M. H., 1884. Traité élémentaire de méchanique céleste. Gauthier-Villars, Paris.

Sagitov, M. U., 1969. Postoyannaya tyagoteniya i massa Zemli. Nauka, Moskva. (Gravitational Constant and the Mass of the Earth; in Russian).

Stegena, L., and Sagitov, M. U. (Eds.), 1979. The Constant of Gravitation. Akademiai Kiadó, Budapest.

Stokes, G. G., 1849. On the variation of gravity at the surface of the earth. Trans. Cambridge Phil. Soc., vol. **VIII**, p. 672.

Švancara, J., 1986. Kvantitativní interpretacní metody v uzité gravimetrii. Geofyzika, Brno. (Quantitative Interpretation Methods in Applied Geophysics; in Czech, not published).

Talwani, M., 1973. Computer usage in the computation of gravity anomalies. In: Methods in Computational Physics, Vol **13** (ed B. A. Bolt), 343-389. Academic Press, New York.

Talwani, M., et al., 1959. Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone. J. Geophys. Res., **64**, 49-59.

Trkal, V., 1956. Mechanika hmotných bodu a tuhého telesa. NCSAV, Praha. (Mechanics of Particles and of a Rigid Body; in Czech).

Vening Meinesz, F. A., 1928. A formula expressing the deflection of the plumb-line in the gravity anomalies and some formulae for the gravity-field and the gravity-potential outside the geoid. Proc. Sect. Sci. Kon. Akad. v. Wet., **31**, Amsterdam, 315.

Vondrák, J., 1983. Co je to projekt MERIT? Říše hvezd, **64**, 9-12. (What is the MERIT project?; in Czech).

Wahr, J., 1996. Geodesy and Gravity. (Class Notes). Samizdat Press, Boulder, CO.