Torge · Geodesy

Wolfgang Torge

Geodesy

Third completely revised and extended edition

W DE G

Walter de Gruyter · Berlin · New York 2001

Author
Dr.-Ing. Wolfgang Torge
Institut für Erdmessung
Universität Hannover
Schneiderberg 50
30167 Hannover
Germany

In memory of

Hans-Georg Wenzel

1945 – 1999

Preface to the Third Edition

The *first edition* of this book was published in 1980 as an English translation of the book "Geodäsie," which was printed in the German language in 1975 within the "Göschen Series" of Walter de Gruyter and Co., Berlin and New York. A thorough revision of the original text was undertaken at that time, but the fundamental structure of the book was retained. The *second edition* (1991) added the remarkable development of space techniques and their effects on positioning and gravity field determination. Though it represented an extension of the first edition, the basic subdivision into six chapters was not altered.

In contrast to the relatively minor changes of the second edition, the *present edition* has been completely revised and significantly extended. This was necessary in order to account for the tremendous changes that geodesy, on the whole, has experienced over the past thirty years, driven by the progress in space techniques and data acquisition and evaluation in general. A particular consequence is that geodesy has transitioned to a three-dimensional concept, based on a global reference system, with far-reaching consequences for geodetic practice. High data rates and improved accuracy require a more rigorous consideration of time as a fourth dimension and has led to a growing contribution to geodynamics research.

The significant extension of the third edition is demonstrated by increases in the number of pages (416 versus 264), the number of figures (184 versus 137), and the number of references. The references increased by about 50 percent to more than 700, half of which are new entries. There are now eight chapters instead of six, and their content more clearly mirrors the fundamental changes of geodesy.

The *"Introduction"* still contains the definition and the over 2000 years history of geodesy. It now also includes the three and four-dimensional geodetic concepts. The survey of geodetic organizations and literature has been updated as well. The new chapter on *"Reference Systems"* comprises the fundamentals of geodetic reference and time systems and their realization and mutual transformation. This information was previously dispersed throughout the book. The third chapter, *"The Gravity Field of the Earth,"* is similar to the corresponding chapter in the previous edition, with some additions to the description of the gravity field geometry and the spherical harmonic expansion. The introduction to the geoid is now included here, while the earth tides section has been moved to the last chapter. The next chapter, *"The Geodetic Earth Model,"* was extracted from the chapter on geodetic reference systems in the previous edition. A separate treatment appeared necessary, especially due to the

importance of the "Geodetic Reference System 1980" and the "World Geodetic System 1984."

The chapters on measurement and evaluation methods comprise the core of the book and have been completely reorganized and revised. The chapter *"Methods of Measurements"* now starts with a detailed description of atmospheric refraction, which affects all geometric methods of geodesy. The dominating role of satellite geodesy is recognized by treating this subject early. Space for classical methods such as Doppler positioning has been reduced, while the Global Positioning System (GPS) has naturally received increased emphasis (now with 10 pages compared to about 4 in the previous edition). Developing techniques such as satellite-to-satellite tracking and satellite gravity gradiometry have been given separate space. Very Long Baseline Interferometry is now included in the section on geodetic astronomy. Gravimetry was updated with respect to absolute and airborne techniques. Terrestrial geodetic measurements concentrate on combined angle and distance observations over shorter distances and on precise leveling. Discussion of tilt and strain measurements are also found here. The former subdivision of *"Methods of Evaluation"* into astrogeodetic, gravimetric, satellite and combined methods has been replaced by an introductory section on the residual gravity field, two large blocks on positioning and gravity field modeling, and a section on combination solutions. Positioning starts from the three-dimensional model, followed by horizontal positioning and height determination, after proper reductions to the ellipsoid and the geoid, respectively. The effect of topography on gravity field modeling is now discussed in more detail, and a clear distinction is made between global and local modeling, where the astrogeodetic methods have also been included. The combined methods comprise the functional approach by earth models and the statistical approach through least squares collocation.

The seventh and the eighth chapters reflect the effects of the developments in geodesy on national and continental networks and on the geosciences. The chapter on *"Geodetic and Gravimetric Networks"* is now free of computational formulas and concentrates on the design and establishment of networks, with special emphasis on the present transition to three-dimensional networks and absolute gravimetry. The final chapter, *"Structure and Dynamics of the Earth,"* has been extracted from the previous edition's chapter on "Methods of Evaluation, Global Geodesy." This chapter considers the increasing role which geodesy plays within the geosciences and, consequently, includes an introduction to the geophysical earth model, especially to the upper layers of the earth. Geodetic contributions to geodynamics research are now described more systematically and extensive, ranging from earth rotation over sea level variations, crustal movements, and gravity variations to earth tides.

The text is again illustrated by a large number of figures. Many of the figures depict fundamental geodetic relations, while others show examples of recent instrumentation and geodetic data processing results. The latter case required a thorough revision of the figures and a revision of the associated passages of the text. Among the new examples for geodetic instruments we have geodetic GPS receivers, altimeter satellites, absolute and relative gravimeters, total stations, and digital levels. The results include the International Celestial and Terrestrial Reference Frames, The World Geodetic System 1984 (status 2000), the geopotential model EGM96, the gravimetric geoids for the U.S.A. and for Europe, and the classical and modern networks in those regions, as well as examples of recent geodetic contributions to the investigation of geodynamics.

The primary purpose of the book is to serve as a basic textbook oriented towards students of geodesy, geophysics, surveying engineering, cartography, and geomatics, as well as students of terrestrial and space navigation. The book is also a valuable reference for geoscientists and engineers facing geodetic problems in their professional tasks.

The contents of the book are based in part on lecture courses given at the University of Hannover, Germany and on guest lectures given abroad. The author is indebted to individuals and organizations for providing illustrations; due credit is given in the figure captions. He thanks M.Sc. Kyle Snow, who checked the English text with extreme care, included the formulas and figures, and prepared the manuscript for printing. The majority of the figures have been drawn by cand. geod. Anke Daubner, Dipl.-Ing. Andreas Lindau handled the electronic mailing and text storage, and assisted in the final proof-reading. The staff of the Institut für Erdmessung assisted in manifold ways in the preparation of the manuscript. All this help is gratefully acknowledged. The outstanding cooperation of the publisher, proved over a nearly 30 years association, continued and calls for a cordial thanks to Dr. Manfred Karbe and the staff at Walter de Gruyter.

Hannover, February 2001   Wolfgang Torge

Contents

1 Introduction

1.1 Definition of Geodesy

According to the classical definition of Friedrich Robert Helmert (1880), "*geodesy* (γη = earth, δαιω = I divide) *is the science of the measurement and mapping of the earth's surface.*" Helmert's definition is fundamental to geodesy even today. The surface of the earth, to a large extent, is shaped by the earth's gravity, and most geodetic observations are referenced to the earth's gravity field. Consequently, the above definition of geodesy includes the determination of the earth's *external gravity field*. The original focus of geodesy has expanded to include applications in ocean and space exploration. For example, Geodesy, in collaboration with other sciences, now includes the determination of the *ocean floor* and the surfaces and gravity fields of *other celestial bodies*, such as the moon (lunar geodesy) and planets (planetary geodesy). Finally, the classical definition has to be extended to include *temporal variations* of the earth's surface and its gravity field.

With this extended definition, geodesy is part of the *geosciences* and *engineering sciences*, including navigation and geomatics (e.g., NATIONAL ACADEMY OF SCIENCES 1978). Geodesy may be divided into the areas of global geodesy, geodetic surveys (national and supranational), and plane surveying. *Global geodesy* includes the determination of the shape and size of the earth, it's orientation in space, and it's external gravity field. A *geodetic survey* is for the determination of the earth's surface and gravity field over a region that typically spans a country or a group of countries. The earth's curvature and gravity field must be considered in geodetic surveys. In *plane surveying* (topographic surveying, cadastral surveying, engineering surveying), the details of the earth's surface are determined on a local level, and thus curvature and gravity effects are generally ignored.

There is a close relation between global geodesy, geodetic surveying, and plane surveying. Geodetic surveys are linked to reference frames (networks) established by global geodesy; these surveys adopt the global parameters for the figure of the earth and its gravity field. On the other hand, the results of geodetic surveys may contribute to the work of the global geodesist. Plane surveys, in turn, are generally referenced to control points established by geodetic surveys. Plane surveys are used extensively in the development of national and state map-series, cadastral information systems, and in civil engineering projects. The measurement and data evaluation methods used in national geodetic surveys are often similar to those used in global geodetic work. For instance, space methods (satellite geodesy) have long been a dominant technique in global geodesy and

are now commonly used in regional and local surveys. This requires a more detailed knowledge of the gravity field at the regional and local scales.

With the corresponding classification in the English and French languages, the concept of "*geodesy*" (la géodésie, "höhere Geodäsie" after *Helmert*) in this text refers only to global geodesy and geodetic surveying. The concept of "*surveying*" (la topométrie, Vermessungskunde or "niedere Geodäsie" after Helmert) shall encompass plane surveying.

In this volume, geodesy is treated only in the more restrictive sense as explained above (excluding plane surveying), and limited to the planet earth. Among the numerous textbooks on surveying we mention KAHMEN and FAIG (1988), KAHMEN (1997), ANDERSON and MIKHAIL (1998), BANNISTER et al. (1998). For lunar and planetary geodesy see BILLS and SYNNOTT (1987), KAULA (1993), and NEREM (1995a); numerical values of astrometric and geodetic parameters are given by Yoder (1995).

1.2   The Problem of Geodesy

Based on the concept of geodesy defined in [1.1], the problem of geodesy may be described as follows:

"The problem of geodesy is to determine the figure and external gravity field of the earth and of other celestial bodies as a function of time, from observations on and exterior to the surfaces of these bodies."

This *geodetic boundary-value problem* incorporates a geometric (figure of the earth) and a physical (gravity field) formulation; both are closely related.

By the *figure of the earth* we mean the physical and the mathematical surface of the earth as well as a geodetic reference earth, e.g., MORITZ (1990).

The *physical surface* of the earth is the border between the solid or fluid masses and the atmosphere. The *ocean floor* is included in this formulation, being the bounding surface between the solid terrestrial body and the oceanic water masses. The irregular surface of the solid earth (continental and ocean floor topography) cannot be represented by a simple mathematical (analytical) function. It is therefore described point wise by the *coordinates of control points*. Given an adequately dense control network, the detailed structure of this surface can be determined by interpolation of data from terrestrial topographic and hydrographic surveying (KRAUS 1992/97, KRAUS and SCHNEIDER 1988/90, HAKE and GRÜNREICH 1994, KAHMEN 1997). On the other hand, the *ocean surfaces* (70% of the earth's surface) are easier to represent. If we neglect the effects of ocean

currents and other "disturbances", the ocean surfaces form a part of a level or equipotential surface of the earth's gravity field (surface of constant gravity potential). We may think of this surface as being extended under the continents and identify it as the *mathematical figure of the earth*, which can be described by a condition of equilibrium (HELMERT 1880/1884). J.B. LISTING (1873) designated this level surface as *geoid*.

The great mathematician and geodesist *Carl Friedrich Gauss* (1777 – 1855) had already referred to this surface: "What we call surface of the earth in the geometrical sense is nothing more than that surface which intersects everywhere the direction of gravity at right angles, and part of which coincides with the surface of the oceans" (C.F. Gauss: "Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen und Altona," Göttingen 1828), see also MORITZ (1977).

The description of the geoid's properties is the physical aspect of the problem of geodesy. In solving this problem, the earth's surface and the geoid are considered as bounding surfaces in the earth's gravity field, the field to which geodetic observations are referenced. Based on the law of gravitation and the centrifugal force (due to the earth's rotation), the *external gravity field* of the earth, or any other celestial body, can be modeled analytically and described by a large number of model parameters. A geometric description is given by the infinite number of *level surfaces* extending completely or partially (as the geoid) exterior to the earth's surface.
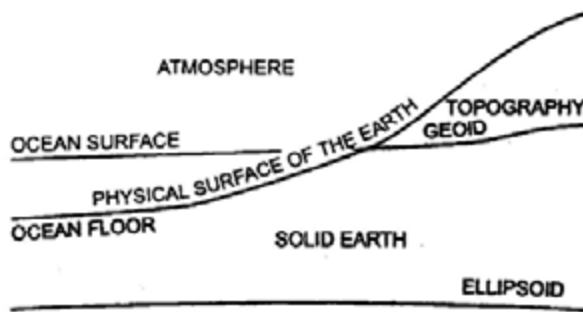
Fig. 1.1. Physical surface of the earth, geoid, and ellipsoid

Reference systems are introduced in order to describe the orientation of the earth and other bodies in space (celestial reference system) as well as their surface geometry and gravity fields (terrestrial reference system). The definition and realization of these systems has become a major part of global geodesy; the use of three-dimensional Cartesian coordinates in Euclidean space is adequate in this context. However, due to the demands of practitioners, *reference surfaces* are introduced to distinguish between curvilinear surface coordinates for horizontal positioning and heights above some zero-height surface for vertical positioning.

Because of its simple equation, an *ellipsoid* of rotation, flattened at the poles, is well suited for describing horizontal positions, and consequently is used as a reference surface in geodetic surveying. In plane surveying, the *horizontal plane* is generally a sufficient reference surface. Because of its physical meaning, the geoid (or any other reference defined in the earth's gravity field) is well suited as a reference surface for heights. For many applications, a *geodetic reference earth* (earth model, normal earth) is needed. It is realized through a *mean-earth ellipsoid* that optimally approximates the geometry (geoid) and the gravity field of the earth. Fig. 1.1 shows the mutual location of the surfaces to be determined in geodesy.

The body of the earth and its gravity field are subject to *temporal variations* of secular, periodic, and abrupt nature, which can occur globally, regionally, and locally. These variations also influence the orientation of the earth. Modern geodetic measurement and evaluation techniques are used to detect these variations to a high level of accuracy. If time-independent results are required, geodetic observations must be corrected for temporal variations. By determining temporal variations, the science of geodesy contributes to the investigation of the kinematic and dynamic properties of the terrestrial body. Accordingly, the figure of the earth and the external gravity field must be considered as time dependent quantities: "Four-dimensional geodesy" (MATHER 1973).

1.3   Historical Development of Geodesy

The formulation of the problem of geodesy as described in [1.2] did not fully mature until the nineteenth century. However, the question of the figure of the earth was contemplated in antiquity. In fact, geodesy together with astronomy and geography are among the oldest sciences dealing with the planet earth. Superseding the use of the *sphere* as a model for the earth [1.3.1], the oblate *rotational ellipsoid* became widely accepted as the model of choice in the first half of the eighteenth century [1.3.2]. The significance of the *gravity field* was also recognized in the nineteenth century, leading to the introduction of the geoid [1.3.3]. In the second half of the twentieth century, satellite techniques permitted the realization of the *three-dimensional* concept of geodesy [1.3.4]. A drastic increase in the accuracy of geodetic observations required that time variations be taken into account. This led to the concept of *four-dimensional* geodesy [1.3.5]. Extensive material on geodetic history is found in TODHUNTER (1873), FISCHER (1975), BIALAS (1982), SMITH (1986) and LEVALLOIS (1988).

1.3.1   The Spherical Earth Model

Various opinions on the figure of the earth prevailed in the past, e.g., the notion of an *earth disk* encircled by oceans (*Homer's* Iliad ~ 800 B.C., *Thales of Milet* ~

600 B.C.). Considering the sphere aesthetically appealing, *Pythagoras* (~ 580 – 500 B.C.) and his school proposed a spherical shaped earth. By the time of *Aristotle* (384 – 322 B.C.), the spherical concept was generally accepted and even substantiated by observations. For example, observers noted the round shadow of the earth in lunar eclipses and the apparent rising of an approaching ship on the horizon. In China the spherical shape of the earth was recognized in the first century A.D.
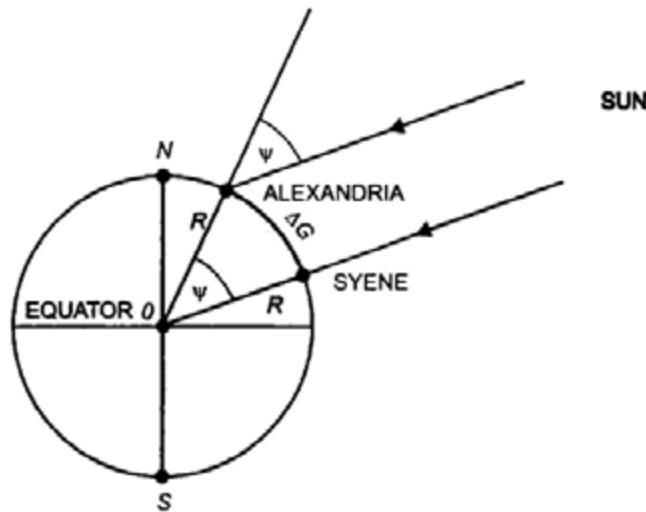
Fig. 1.2. Arc measurement of Eratosthenes

The founder of geodesy is *Eratosthenes* of Alexandria (276 – 195 B.C.), who, based on the assumption of a spherical earth, deduced the earth radius from measurements (SCHWARZ 1975). The principle of the *arc-measurement* method developed by Eratosthenes was still applied in modern time: from geodetic measurements, the length ΔG of a meridian arc is determined; astronomical observations furnish the associated central angle ψ (Fig. 1.2). The radius of the earth is then given by

$$R = \frac{\Delta G}{\psi}.$$
(1.1)

Eratosthenes found that at the time of the summer solstice the rays of the sun descended vertically into a well in Syene (modern day Assuan). Whereas in Alexandria (approximately on the same meridian as Syene), the sun's rays formed an angle with the direction of the plumb line. From the length of the shadow of a vertical staff ("gnomon") produced in a hemispherical shell ("skaphe"), Eratosthenes determined this angle as 1/50 of a complete circle, i.e., $\psi = 7°12'$.

From Egyptian cadastre maps, which were based on the information of "bematists" (step counters), Eratosthenes estimated the distance from Syene to Alexandria to be 5000 stadia. With the length of an Egyptian stadium assumed as 148.5 m, the earth radius is computed to be 5909 km from Eratosthenes estimations. This value departs from the radius of a mean spherical earth (6371 km) by only -7%. Another ancient determination of the earth's radius is attributed to *Posidonius* (135 – 51 B.C.). Using the meridian arc from Alexandria to Rhodes, he observed the star Canopus to be on the horizon at Rhodes, while at a culmination height of 7°30′ at Alexandria, which again equals the central angle between the cities.

During the middle ages in Europe, the question of the figure of the earth was not pursued further. Documentation from China shows that an astronomic-geodetic survey between 17° and 40° latitude was carried out by the astronomers *Nankung Yüeh* and *I-Hsing* around 725 A.D. in order to determine the length of a meridian. A 1° arc length was measured directly with ropes by the Arabs (~ 827 A.D.) northwest of Bagdad by the caliphate of *Al-Mámûn*. At the beginning of the modern age, the French physician *Fernel* (1525) observed the geographical latitudes of Paris and Amiens using a quadrant. He computed the corresponding surface distance from the number of rotations of a wagon wheel.

Later arc measurements based on the spherical earth model benefited from fundamental advances in instrumentation technology (1611, *Kepler* telescope) and methodology. After the initial application of *triangulation* by *Gemma Frisius* (1508 – 1555) in the Netherlands, and by *Tycho Brahe* (1546 – 1601) in Denmark, the Dutchman *Willebrord Snellius* (1580 – 1626) conducted the first triangulation to determine the figure of the earth, HAASBROECK (1968). Snellius used triangulation to measure the arc between Bergen op Zoom and Alkmaar (Holland). The hitherto inaccurate arc length estimate by direct measurement was replaced by an indirect procedure. The angles of a triangulation network were observed with high precision instruments, and the scale was accurately determined by precise measurement of short baselines. With proper reduction of the observations to the meridian, the length of arc could be accurately calculated. A direct length measurement using a chain was employed by *Norwood* when determining the meridian arc between London and York (1633 – 1635).

The method of *reciprocal zenith angles* is yet another technique that has been used to determine the central angle between points on a meridian arc. The Italian fathers *Grimaldi* and *Riccioli* used this method in 1645 between Bologna and Modena (Fig. 1.3). The central angle may be computed from the zenith angles z_1 and z_2 observed at locations P_1 and P_2 according to

$$\psi = z_1 + z_2 - \pi .$$

(1.2)
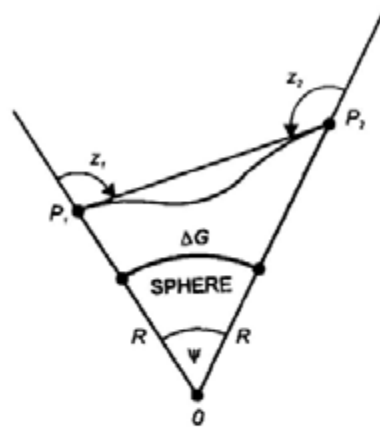
Fig 1.3. Central angle and reciprocal zenith angles

This procedure does not yield satisfactory results due to the inaccurate determination of the curvature of light rays (refraction anomalies).

Through the initiative of the Academy of Sciences (founded in Paris, 1666), France assumed the leading role in geodesy in the seventeenth and eighteenth centuries. In 1669 – 1670 the French abbey *J. Picard* measured the meridian arc through Paris between Malvoisine and Amiens with the aid of a triangulation network; he was the first to use a telescope with cross hairs. The value Picard obtained for the radius of the earth (deviation of +0.01%) aided *Newton* in the verification of the law of gravitation, which he had formulated in 1665 – 1666.

1.3.2   The Ellipsoidal Earth Model

In the sixteenth and seventeenth centuries, new observations and ideas from astronomy and physics decisively influenced the perception of the figure of the earth and its position in space. N. *Copernicus* (1473 – 1543) achieved the transition from the *geocentric* universe of *Ptolemy* to a *heliocentric* system (1543: "De revolutionibus orbium coelestium"), which *Aristarchus of Samos* (~ 310 – 250 B.C.) had already postulated. *J. Kepler* (1571 – 1630) discovered the laws of planetary motion (1609: "Astronomia nova...", 1619: "Harmonices mundi"), and *Galileo Galilei* (1564 – 1642) established the fundamentals for mechanical dynamics (law of falling bodies and law of pendulum motion).

In 1666, the astronomer *J. D. Cassini* observed the flattening of the poles of Jupiter. On an expedition to Cayenne to determine martian parallaxes (1672 – 1673), the astronomer *J. Richer* discovered that a one-second pendulum regulated in Paris needed to be shortened in order to regain oscillations of one second. From this observation, and on the basis of the law of pendulum motion, one can infer an ncrease in gravity from the equator to the poles. This effect was

confirmed by *E. Halley* when comparing pendulum measurements in St. Helena to those taken in London (1677 – 1678).

Building on these observations and on his theoretical work on gravitation and hydrostatics, *Isaac Newton* (1643 – 1727) developed an earth model based on physical principles: "Philosophiae Naturalis Principia Mathematica" (1687). Based on the law of gravitation, Newton proposed a rotational ellipsoid as an equilibrium figure for a homogeneous, fluid, rotating earth. The flattening

$$f = \frac{a-b}{a} \tag{1.3}$$

(a = semimajor axis, b = semiminor axis of the ellipsoid) of Newton's ellipsoid was 1/230. He also postulated an increase in gravity acceleration from the equator to the poles proportional to $\sin^2 \varphi$ (φ = geographical latitude). At the same time, the Dutch physicist *Christian Huygens* (1629 – 1695), after having developed the principle of pendulum clocks and the law of central motion, calculated an earth model flattened at the poles ("Discours de la Cause de la Pesanteur," 1690). Shifting the source of the earth's attractive forces to the center of the earth, he obtained a rotationally-symmetric equilibrium-surface with a meridian curve of fourth order and flattening of 1/576.

Arc measurements at *various* latitudes were required to verify the proposed ellipsoidal earth-models. Theoretically, the length of a one-degree arc (meridian arc for a difference of 1° in latitude), in the case of flattened poles, should increase pole-ward from the equator. The ellipsoidal parameters a, b or a, f can be computed from two arc measurements.
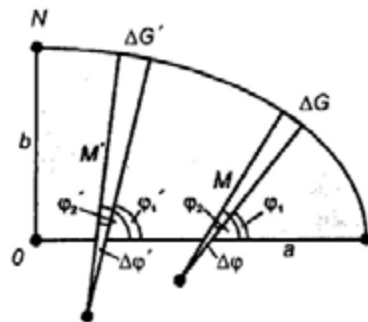
Fig. 1.4. Latitude arc measurement

We distinguish between arc measurements along an ellipsoidal meridian (latitude arc measurement), along a parallel (longitude arc measurement), and arc measurements oblique to the meridian.

For the computations in a *latitude arc-measurement* (Fig. 1.4), the angles $\Delta\varphi = \varphi_2 - \varphi_1$ and $\Delta\varphi' = \varphi_2' - \varphi_1'$ are formed from the observed geographic latitudes $\varphi_1, \varphi_2, \varphi_1'$, and φ_2'. The corresponding meridian arcs ΔG and $\Delta G'$ are obtained from triangulation networks. For short arcs, one can replace the meridian ellipse by the osculating circle having the meridian radius of curvature $M = M(\varphi)$ evaluated at the mean latitude $\varphi = (1/2)(\varphi_1 + \varphi_2)$, where M is also a function of the ellipsoidal parameters a, f. From $\Delta G = M\Delta\varphi$ and $\Delta G' = M'\Delta\varphi'$, a and f may be determined. The larger the latitude interval $\varphi' - \varphi$, the more accurate the computed flattening; whereas, the accuracy of the semi-major axis length a depends in particular on the lengths of the meridian arcs.

For *longitude arc measurements*, corresponding relations are used between the arc lengths measured along the parallels and the difference of the geographical longitudes observed at the end points of the arcs. Arc measurements *oblique to the meridian* require a proper azimuth determination for reduction to the meridian.

Initial evaluations of the older arc measurements (*Snellius, Picard,* among others) led to an earth model elongated at the poles. The same result was obtained by *La Hire, J. D.* and *J. Cassini*. They extended the arc of Picard north to Dunkirk and south to Collioure (1683 – 1718), with a latitude difference of $8°20'$. The computations from two arc segments yielded a "negative" flattening of $-1/95$, which can be attributed primarily to measurement errors in the astronomic latitudes. The intense dispute between the supporters of Newton and those of the Cassinis over the figure of the earth was resolved by two further arc measurement campaigns sponsored by the French Academy of Sciences.

Maupertuis and *Clairaut*, among others, participated in the expedition to Lapland (1736 – 1737). The results of the Lapland arc measurement (average latitude $66°20'$ and latitude interval $57'.5$) confirmed the polar flattening. Using the arc measurement of the meridian through Paris (revised by *Cassini de Thury* and *La Caille*, 1739 – 1740) the flattening was computed as $1/304$. On a second expedition (1735 – 1744) to Peru (modern day Ecuador), an arc at an average latitude $1°31'$ S and with $3°7'$ amplitude was determined by *Bouguer, La Condamine* and *Godin*. Combining this information with the Lapland arc led to a flattening of $1/210$. The flattening of the earth at the poles was thereby demonstrated by *geodetic* measurements.

A *synthesis* between the physical and the geodetic evidence of the ellipsoidal shape of the earth was finally achieved by *A. C. Clairaut* (1713 – 1765). The theorem (1743), which bears his name, permits the computation of the flattening

from two gravity measurements at different latitudes, cf. [4.2.2]. A first application of Clairaut's theorem is due to *P. S. Laplace* (1799), who from only 15 gravity values derived a flattening of $1/330$. The wider application of this *"gravimetric method"* suffered from the lack of accurate and well distributed gravity measurements and from the difficulty of reducing the data to the earth ellipsoid. Such problems were not overcome until the twentieth century.

With the rotational ellipsoid commonly accepted as a model for the earth, numerous *arc measurements* were conducted up to the twentieth century. These measurements generally served as a basis for geodetic surveys, see [1.3.3]. The meridian arc through Paris was extended by *Cassini de Thury* and was included in the first triangulation of France (1733 – 1750). A geodetic connection between the astronomical observatories in Paris and Greenwich (1784 – 1787) was the beginning of the national survey of Great Britain, with the final extension of the Paris meridian arc to the Shetland islands. Particular significance was attained by a measurement on the meridian through Paris, between Barcelona and Dunkirk (1792 – 1798), commissioned by the French National Assembly and carried out by *Delambre* und *Méchain*. The results served for the definition of the meter as a natural unit of length (1799). These observations, combined with the Peruvian arc measurement, yielded an ellipsoidal flattening of $1/334$.

1.3.3   The Geoid, Arc Measurements and National Geodetic Surveys

As recognized by P. S. *Laplace* (1802), *C. F. Gauss* (1828), *F. W. Bessel* (1837), and others, the assumption of an ellipsoidal-earth model is no longer tenable at a high level of accuracy. The deviation of the physical plumb line, to which the measurements refer, from the ellipsoidal normal then can no longer be ignored. This deviation is known as the *deflection of the vertical*. While adjusting several arc measurements for determination of the ellipsoidal parameters, contradictions were found which greatly exceeded the observational accuracy. An initial adjustment of arc measurements was carried out in 1806 by *A. M. Legendre* in his treatise "Sur la méthode des moindres carrées". *C. F. Gauss* was the first to adjust a triangulation network (in and around Brunswick, 1803 – 1807) by the method of least squares (GERARDY 1977).

This led to the improved definition of the "figure of the earth" by *Gauss* and *Bessel*, who clearly distinguished between the physical surface of the earth, the geoid as the mathematical surface, and the ellipsoid as a reference surface approximating it, cf. [1.2]. With the definition of geodesy [1.1], *F. R. Helmert* made the transition to the actual concept of the figure of the earth (MORITZ 1990).

Friedrich Robert Helmert (1843 – 1917), one of the most distinguished geodesists of modern times, was professor of geodesy at the technical university at Aachen and later director of the Prussian Geodetic Institute in Potsdam and of the Central Bureau of the "Internationale Erdmessung". Through his work, geodesy has experienced decisive impulses, the effects of which are still felt

today. In his fundamental monograph (1880/1884), *Helmert* established geodesy as a proper science (WOLF 1993, HARNISCH and HARNISCH 1993).

Despite the discrepancies found from the adjustments, arc measurements continued to be used to determine the dimensions of the earth ellipsoid. However, the deflections of the vertical, being a physically phenomena and hence having systematic characteristics, were treated as random observational errors. As a consequence, this method provided parameters for *best-fitting* ellipsoids, approximating the geoid in the area of the triangulation chains. Many of these ellipsoids have been introduced as "*conventional*" ellipsoids for calculating the national geodetic surveys, and thus arc measurements increasingly became part of the geodetic surveys. Established by triangulation, these national surveys provided control points for mapping, which remain the basis for many national geodetic reference systems today (TORGE 1997). Gravity observations were carried out at most arc measurements and at dedicated campaigns, especially after the foundation of the "Mitteleuropäische Gradmessung", cf. [1.4.2].

We mention the historically important arc of *Gauss* (arc measurement between Göttingen and Altona 1821 – 1824, invention of the Heliotrope, adjustment according to the method of least squares) and its extension to the triangulation of the kingdom of Hannover (until 1844). Initiated by the Danish astronomer *H. Chr. Schumacher*, this arc should become part of a central European meridian arc, running from Denmark to Bavaria (triangulation by *J. G. Soldner*, 1808 – 1828) and further south. *Bessel* and *Baeyer* carried out an arc measurement oblique to the meridian in East Prussia (1831 – 1838), which connected the Russian triangulations (*W. Struve, C. Tenner*) with the Prussian and Danish networks and finally with the French-British arc along the meridian of Paris. Some *long arcs* linking national triangulation-chains were built up over a 100 year period. Some of these were not completed until 1950's, while others were never finished, owing to the replacement of classical geodetic observation techniques by satellite surveying methods. These long arcs include the American meridian arc (Alaska – Tierra del Fuego), the North American longitude arc along the 39° parallel between the Atlantic and the Pacific Oceans, the West European-African arc along the meridian of Paris (Shetlands – Algeria), the Arctic Ocean to Mediterranean Sea meridian arc (Hammerfest – Crete) and the African 30° east meridian arc tied to it (Cairo – Cape Town), the European-Asiatic longitude arc measurements at 48° (Brest – Astrachan) and at 52° latitude (Ireland – Ural Mountains), as well as the latitude and longitude arc measurements in India (*G. Everest, W. Lambdon*).

Since the 1880's, *vertical control networks* were established by geometric leveling within the frame of the national geodetic surveys but *independently* from the horizontal control systems. Heights were referred to a level surface close to the geoid and defined by the mean sea level as observed at a tide gauge. The

geoid was not needed in this separate treatment of horizontal position and height, but it played a major role as a geometric representation of the earth's gravity field.

An inevitable presupposition for the evaluation of large-scale arc measurements, triangulations, and leveling was the regulation of the *measuring units*. It was nearly one century after the introduction of the *meter* in France that a large number of countries agreed upon a definition for the meter at the International Meter Convention in Paris in 1875. Following a recommendation of the "Europäische Gradmessung" in 1883, the International Meridian Conference (Washington, D. C., 1884) adopted the Greenwich meridian as the initial meridian for longitude and the universal day (mean solar day) as the *time* unit referenced to the meridian.

1.3.4   Three-dimensional Geodesy

The three-dimensional concept of geodesy consists of the common treatment of horizontal and vertical positioning within the same mathematical model. This was suggested already by BRUNS (1878), who proposed to determine the surface of the earth pointwise using a spatial polyhedron together with all exterior level surfaces. However, three-dimensional computations were not carried out in practice due to the uncertainty of trigonometric derived height differences over large distances and due to the absence of geoid heights above the ellipsoid, which were required for the proper treatment of geometric leveling.

The concept of three-dimensional geodesy was revived by MARUSSI (1949) and HOTINE (1969), while in 1945 Molodenski demonstrated that the physical surface of the earth and its external gravity field can be determined from surface measurements only, without needing the geoid, MOLODENSKI (1958).

VÄISÄLÄ (1946) introduced *Stellar triangulation* from high altitude balloons as a first step to realize the three-dimensional concept. This technique was followed by *electromagnetic distance measurements* in the 1950's and 1960's, using both terrestrial and airborne methods. Satellite geodesy provided a technological breakthrough after the launch of the Russian satellite Sputnik I in 1957. Observations to orbiting satellites were used to establish control points in a three-dimensional system, and provided valuable gravity field information. Beginning in the 1980's, the NAVSTAR *Global Positioning System (GPS)* now dominates geodetic measuring techniques. The practical problems of geodesy today include the connection of classical horizontal and vertical control networks to the global reference system established by space methods. A particular problem is the determination of the geoid with respect to a global reference ellipsoid.

Recently, *kinematic methods* have gained great importance, especially with the use of GPS. The measuring systems, or their components, are carried on moving

platforms (e.g., satellite, airplane, ship, car) and, by continuous positioning (navigation), provide data referring to the geodetic reference system.

1.3.5 Four-Dimensional Geodesy

The beginning of four-dimensional geodesy may be reckoned from the detection of *polar motion* by *F. Küstner* (1884 – 1885) and first observations of the *earth tides* by *E.v. Rebeur-Paschwitz* (1889 – 1893). Monitoring of *crustal deformations* related to seismic activities began in Japan and the U.S.A. about 100 years ago. Interest in this phenomena was motivated by powerful seismic events, such as the San Francisco earthquake of 1906. In Fennoscandia, precise leveling and tide gauge registrations began in the 1880's and were used to determine the region's large-scale *vertical uplift* caused by postglacial rebound.

Today, the variations of the *earth's rotation* and the movements of the *tectonic plates* are regularly observed through global networks, and a number of regional control networks have been set up, especially at tectonic plate boundaries. *Gravity variations* with time are derived from the analysis of satellite orbits (large-scale) and terrestrial networks (small-scale). The *earth tides* have also been modeled successfully using terrestrial and satellite methods.

Large efforts continue worldwide to measure and analyze all types of geodynamic phenomena by geodetic methods, for instance that described in the NASA (1983) Geodynamics Program, see also MUELLER and ZERBINI (1989).

In the future, geodetic observations will experience a further increase of accuracy and an increase in resolution in space and time. Longer observation series will permit the detection of long-term changes of the earth and its gravity field. Global geodesy and geodetic surveys must take these temporal changes into account. With improved modeling, the four-dimensional aspect of geodesy will increasingly be employed in the evaluation and presentation of geodetic products (LAMBECK 1988, BRUNNER and RIZOS 1990).

1.4 Organization of Geodesy, Literature

1.4.1 National Organizations

The problems of *global geodesy* may be solved only with the international cooperation of institutions involved in nation-wide geodetic work, together with international geodetic services, cf. [1.4.2]. *University* institutes and departments (e.g., geodesy, geophysics, astronomy, and space sciences) pursue fundamental *research*. In some countries, *academy* or *governmental* institutes are also engaged in geodetic research (China: Institute of Geodesy and Geophysics, Wuhan;

Finland: Finnish Geodetic Institute; Germany: Deutsches Geodätisches Forschungsinstitut, München, Geoforschungszentrum Potsdam; Russia: Institute of Physics of the Earth, Moscow). The *geodetic surveys* are carried out according to the guidelines of the official survey system, either by central agencies or by decentralized institutions (Australia: Australian Surveying and Land Information Group; Canada: Geodetic Survey Division, National Resources Canada, China: National Bureau of Surveying and Mapping; France: Institut Géographique National; Germany: State geodetic surveys in cooperation with the federal Bundesamt für Kartographie und Geodäsie BKG; Great Britain: Ordnance Survey; India: Survey of India; Japan: Geographical Survey Institute; Russia: Federal Service of Geodesy and Cartography; South Africa: Surveys and Land Information; U.S.A.: National Geodetic Survey, National Oceanic and Atmospheric Administration NOAA, (formerly U.S. Coast and Geodetic Survey).

In addition to these, a number of non-geodetic institutions, in the course of their special tasks and projects, are also concerned with geodetic problems. These groups develop theory, measuring systems and methods, and in particular are involved with the collection and evaluation of geodetic data. We mention *space agencies* (e.g., Goddard Space Flight Center of NASA, Greenbelt, Md.; Centre National d'Etudes Spatiales, Toulouse), *geologic* and *hydrographic services* (China: State Seismological Bureau; France: Bureau des Recherches Géographiques et Minières; Germany: Bundesanstalt für Geowissenschaften und Rohstoffe, Bundesamt für Seeschiffahrt und Hydrographie; Great Britain: Institute of Geological Sciences, Institute of Oceanographic Sciences; U.S.A.: U.S. Geological Survey, U.S. Naval Observatory), *University departments* (e.g., Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Cal.; Lamont Doherty Earth Observatory, Columbia Univ., Palisades, N.Y.) and *military agencies* (e.g., U.S.A.: National Imagery and Mapping Agency NIMA, formerly Defense Mapping Agency). More details may be found in Journal of Geodesy 74 (2000): 142 – 154.

1.4.2   International Collaboration

At the beginning of the arc measurements in the kingdom of Hanover (1821), *C. F. Gauss* had already expressed his desire for international collaboration. According to Gauss, this geodetic network would be connected to neighboring triangulation networks, aiming toward an eventual merger of the European observatories. Organized international collaboration originated with the memorandum by the Prussian general *J. J. Baeyer* (1794 – 1885): "Über die Größe und Figur der Erde, eine Denkschrift zur Begründung einer Mitteleuropäischen Gradmessung" (1861). In 1862, the "Mitteleuropäische Gradmessung" was founded in Berlin as the first international scientific association of significance; Baeyer became its first president. After expanding to the "Europäische Gradmessung" (1867) and to the "Internationale Erdmessung" ("Association Géodésique Internationale," 1886), the association engaged in fruitful activity,

which was especially inspired by the works of *Helmert* as director of the Central Bureau (LEVALLOIS 1980, TORGE 1996).

After the dissolution of the "Internationale Erdmessung" during the first World War, the "*International Union of Geodesy and Geophysics*" (IUGG) was founded in 1919. Today this organization has a membership of 75 countries. It consists of one geodetic and six geophysical associations. The "*International Association of Geodesy*" (IAG) is directed by a President who is elected every four years, and who is assisted by a Vice President and a General Secretary. The IUGG and IAG meet at General Assemblies at four-year intervals.   In addition, numerous symposia and scientific conferences are organized to treat special themes; among these are the IAG Scientific Assemblies, which are held between the General Assemblies.

The IAG consists of five sections: Positioning, Advanced Space Technology, Determination of the Gravity Field, General Theory and Methodology, Geodynamics. *Commissions* are established for continuing problems, whereas transient problems are treated by *special study groups*. In addition, the IAG, in collaboration with other scientific organizations, maintains the following permanent services: International GPS Service (IGS) with the Central Bureau at the Jet Propulsion Laboratory, Pasadena, California; Bureau Gravimetrique International (BGI), Toulouse; International Geoid Service (IGeS), Milano; International Center for Earth Tides (ICET), Brussels; International Earth Rotation Service (IERS) with the Central Bureau at the Bundesamt für Kartographie und Geodäsie (BKG), Frankfurt, a.M. (and until December 31, 2000: Observatoire de Paris); Permanent Service for Mean Sea Level, Bidston Observatory, Merseyside, U.K.; Bureau International des Poids et Mésures-Time Section, Sèvres; International Laser Ranging Service (ILRS) and International VLBI Service for Geodesy and Astrometry (IVS), since 1999. The IAG also maintains an information and a bibliographic service and organizes summer schools in many parts of the world (IAG 1997). A restructuring of IAG is under discussion and may lead to major changes (SCHWARZ 2000a).

1.4.3   Literature

References to textbooks for geodesy and related fields (mathematics, physics, astronomy, geophysics, surveying engineering and mapping) will be found in the running text. A list of geodetic and geodetically relevant publication series is given in Journal of Geodesy 74 (2000): 155 – 162.

We mention in particular the Journal of Geodesy (formerly Bulletin Géodésique and Manuscripta Geodaetica, Springer: Berlin-Heidelberg-New York), which is the official journal of the IAG. The results of each General Assembly of the IAG are compiled in the Travaux (Proceedings). National reports are collected and

stored at the Central Bureau of the IAG. The proceedings of IAG symposia are published in a separate series (Springer).

Among the recent *scientific-technical journals* in the field of geodesy, geophysics, navigation, and surveying we mention in particular: Acta Geodaetica, Geophysica et Montanistica Hungarica (Hungary), Acta Geodaetica et Cartographica Sinica (China), Annali di Geofisica (Italy), Artificial Satellites (Poland), Australian Journal of Geodesy, Photogrammetry and Surveying, The Australian Surveyor, Bolletino de Geodesia e Scienze Affini (Italy), Allgemeine Vermessungsnachrichten (Germany), Bolletino die Geofisica Teorica ed Applicata (Italy), EOS Transactions (American Geophysical Union), Geodesia (The Netherlands), Geomatica (Canada), Geodeziya i Aerosyemka, Geodeziya i Kartografiya (Russia), Geophysical Journal International (U.K.), Geophysical Research Letters (U.S.A.,), Geophysics (U.S.A.), GPS-World (U.S.A.), GPS Solutions (U.S.A.), Journal of Earthquake Prediction Research (China/Russia), Journal of Geodynamics (The Netherlands), Journal of the Geodetic Society of Japan, Journal of Geophysical Research (U.S.A.), Journal of Surveying Engineering (U.S.A.), Marine Geodesy (U.S.A.), The Journal of Navigation (U.S.A.), Österreichische Zeitschrift für Vermessungswesen und Geoinformation (Austria), Physics and Chemistry of the Earth A: Solid Earth and Geodesy (The Netherlands),   Reviews of Geophysics and Space Physics (U.S.A.), Revista Cartografica (Mexico), Surveying and Land Information Systems (U.S.A.), Studia Geophysica et Geodaetica (Czech Republic), Survey Review (U.K.), Surveys in Geophysics (The Netherlands), Tectonophysics (The Netherlands), Vermessung, Photogrammetrie und Kulturtechnik (Switzerland), Zeitschrift für Vermessungs-wesen (Germany).

Technical Reports are issued by university and research institutes, as well as by some governmental agencies. We mention here: Bull. d'Inf. Bureau Gravimetrique International, Toulouse; Bull. d'Inf. Marées Terrestres, Brussels; Bull. Earthquake Research Inst., Univ. of Tokyo; Bull. Geograph. Survey Inst., Tokyo; Geod. Geophys. Arb. in der Schweiz, Schweiz. Geod. Komm., Zürich; Geowiss. Mittl. Studieng. Verm.wesen, TU Wien; IERS Techn. Notes, Paris; IGS Techn. Reports JPL, Pasadena, U.S.A.;   Journal of Wuhan Technical University of Surveying and Mapping; Mitt. Bundesamt Kart. u. Geod., Frankfurt a.M.; Mitt. Geod. Inst. Univ. Bonn; Mitt. Geod. Inst. TU Graz; Mitt. Inst. Geod. Photogr. ETH Zürich; NASA Goddard Space Flight Center Rep., Greenbelt, Md.; Nat. Survey and Cadastre, Geod. Div. Techn. Rep., Copenhagen; Nederlandse Comm. voor Geodesie Publ.; NIMA Techn. Rep., Washington D.C.; NOAA-NOS-National Geod. Survey Techn. Rep., Rockville, Md.; Publ./Rep. Finnish Geod. Inst. Helsinki; Publ. Division of Geomatics, Univ. of Calgary; Rep. Dep. of Geodetic Science and Surveying, The Ohio State Univ., Columbus, Ohio; Rep. on Geodesy, Inst. of Geodesy and Geod. Astronomy, Warsaw Univ. of Technology; Math. and Phys. Geodesy, TH Delft; Schriftenreihe d. Institute d. Fachber. Vermessungswesen, Univ. Stuttgart;

Schriftenr. Univ. Studiengang Vermessungswesen, Univ. der Bundeswehr, München; Univ. Rep. School of Geomatic Engineering, Univ. of New South Wales, Sydney; Veröff. Bayer. Komm. für die Internationale Erdmessung, München; Veröff. Deutsche Geod. Komm., München; Wiss. Arb. Fachr. Vermessungswesen, Univ. Hannover.

2 Reference Systems

Reference systems are introduced in order to model geodetic observations as a function of unknown parameters of interest. The coordinate systems are defined in terms of orientation, metrics, and curvature; they are three-dimensional in principle (HEITZ 1988). A fourth dimension, time, enters through the mutual motion of the earth and other celestial bodies and through the earth's deformations. As with the earth, reference systems can be defined for the moon and the planets in the solar system.

Basic units and fundamental constants are foundational to the geodetic measurement and modeling process [2.1]. Time systems are based either on processes of quantum physics or on the daily rotation of the earth [2.2]. Global reference systems are realized through reference frames, e.g., the ITRF, which is established and maintained by the International Earth Rotation Service [2.3]. We distinguish between the space-fixed celestial reference system [2.4] and the earth-fixed terrestrial reference system [2.5] (KOVALEVSKY et al. 1989). In addition, gravity related reference systems have to be introduced, as most geodetic observations refer to the earth's gravity field [2.6].

2.1   Basic Units and Fundamental Constants

Length, mass, and time are basic quantities used in geodesy. The units for these quantities are the meter (m), the kilogram (kg), and the second (s) respectively. They are defined through the International System of Units (Système International d'Unités SI), established in 1960 by the 11th General Conference of Weights and Measures (CGPM) in Paris (MARKOWITZ 1973, BIPM 1991). The definitions are as follows:

- The *meter* is the length of the path traveled by light in vacuum during a time interval of 1/299 792 458 of a second (CGPM 1983).
- The *kilogram* is the unit of mass; it is equal to the mass of the international prototype of the kilogram (CGPM 1901).
- The *second* is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom (CGPM 1967).

The establishment and maintenance of the reference standards for these units is the task of the *Bureau International des Poids et Mésures* (BIPM), located in Sèvres, France. BIPM cooperates with the national laboratories of standards under the guidelines of the International Meter Convention (1875). These national laboratories include the National Institute of Standards and

Technology, Gaithersburg, Md., U.S.A.; the National Physical Laboratory, Teddington, U.K.; and the Physikalisch-Technische Bundesanstalt, Braunschweig, Germany.

The *realization* of the *meter* is based on interferometric measurements (relative uncertainty 10^{-12}) using light with highly stable frequencies (stabilized lasers). The international *kilogram* prototype has been kept in BIPM since 1889; national prototypes are related to it with an uncertainty of 10^{-9}. The BIPM Time Section (until 1987: Bureau International de l'Heure BIH, Paris) defines the *second* (relative uncertainty 10^{-14}) and the atomic time scale, cf. [2.2.1].

Previous definitions of the meter and the second were based on natural measures. The *meter* was intended to be one ten-millionth part of the meridian quadrant passing through Paris. Its length was derived from an arc measurement, cf. [1.3.2], and realized in 1799 by a prototype meter bar called "mètre des archives" (legal meter). Following the International Meter Convention, a more stable version (platinium-iridium bar) was manufactured (international meter). It has been preserved since 1889 at the BIPM. This definition (uncertainty 10^{-7}) was valid until 1960 when, for the first time, the wavelength of a certain spectral line of light became the defining quantity.

Since ancient times, the natural measure for *time* has been the daily rotation of the earth about its axis. The mean solar day, cf. [2.2.2], was determined by astronomical observations, and the second was defined as 1/86400 part of that day. From the 1930's on, it became obvious that this definition was uncertain by about 10^{-7} due to the irregularities of the earth's rotation, cf. [2.5.2].

As a supplementary SI unit, the *radian* (rad) is used for *plane angles*:

* The radian is the plane angle between two radii of a circle subtended by an arc on the circumference having a length equal to the radius.

Geodesy, astronomy, and geography also use the *sexagesimal graduation* with 1 full circle = 360° (degrees), 1° = 60′ (minutes), and 1′ = 60″ (seconds, also arcsec). With 2π rad corresponding to 360°, an angle α is transformed from radian to degree by

$$\alpha(°) = \rho(°)\alpha \, \text{rad}, \quad \rho° = 180°/\pi . \tag{2.1}$$

Among the *fundamental constants* used in geodetic models is the *velocity of light* in a vacuum, which is by definition (1983)

$$c = 299\,792\,458\,\text{ms}^{-1}, \tag{2.2}$$

and the *gravitational constant* (CODATA system of physical constants 1986), which is defined as

$$G = (6.672\ 59 \pm 0.000\ 85) \times 10^{-11} \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}. \tag{2.3}$$

Cavendish carried out the first experimental determination of G in 1798 with a torsion balance. Current work concentrates on increasing the relative accuracy of G to better than 10^{-4}. This includes investigations into dependence of G on material, external influences, distance and direction, as well as non-inverse-square properties of gravitation (GILLIES 1987, FISCHBACH and TALMADGE 1999).

Other units and constants used in geodesy, astronomy, and geophysics will be introduced in the corresponding chapters, see also AHRENS (1995), BURŠA (1995), GROTEN (2000).

2.2   Time Systems

Time plays a fundamental role in geodesy. This is due to the fact that most measurement methods use the signal travel-time of electromagnetic waves for positioning, and that a uniform time scale is also needed in order to model the motion of artificial satellites. On the other hand, a time system is required for describing the relative motion of the earth in the solar system with respect to inertial space and for describing earth deformations due to internal and external forces.

Time systems are defined by the unit for a time interval and by a time epoch. They are based either on the definition of the SI second [2.2.1] or on the diurnal rotation of the earth about its axis [2.2.2]. Fundamental descriptions of time systems are found in MUELLER (1969), MORITZ and MUELLER (1987), SEIDELMANN (1992).

2.2.1   Atomic Time, Dynamical Time

A uniform time-scale of high accuracy is provided by the *International Atomic Time* (Temps Atomique International: TAI). It corresponds to the definition of the SI second, cf. [2.1], which has been made approximately equal to the second of the formerly used ephemeris-time. The latter was defined by the motion of the earth about the sun and determined through long-term astronomic observations. The origin of TAI was chosen so that its epoch (January 1, 1958, 0 h) coincided with the corresponding epoch of Universal Time UT1, cf. [2.2.2]. The TAI day comprises 86400s, and the Julian Century has 36525 TAI days.

TAI is *realized* by a large set (more than 200) of atomic clocks (mostly cesium beam frequency standards providing long-term stability and a few hydrogen

masers) maintained at about 60 laboratories around the world. Clock comparisons are performed at a number of timing centers, employing GPS observations for time links, cf. [5.2.5]. From these local determinations, a weighted mean is calculated at the BIPM Time Section. The relative frequency-stability of TAI is between a few 10^{-15} (over minutes to days) and 10^{-13} (over years). Due to relativistic effects, the readings of the atomic clocks are reduced to a common height reference (SI second "on the geoid").

The motions of celestial bodies and artificial satellites have to be described by a strictly uniform time scale (inertial time). This is provided by a *dynamical time*, which is based on motions of bodies in the solar system. Dynamical time scales refer either to the barycenter of the solar system (Barycentric Dynamic Time TDB) or to the geocenter (Terrestrial Time TT). The TT unit is practically equivalent to TAI, with a constant difference resulting from the epoch definition of TAI:

$$TT = TAI + 32.184s.\qquad(2.4)$$

Dynamical time is used in celestial mechanics with Newton's equations of motion, e.g., as an argument for the astronomical ephemeris of the moon and the sun.

2.2.2   Sidereal and Universal Time

The diurnal rotation of the earth provides a natural measure for time. Corresponding time systems are introduced in order to relate earth-based observations to a space-fixed system: Sidereal and Universal (solar) Time. Hereby, two periodic motions of the earth play a role (Fig. 2.1):
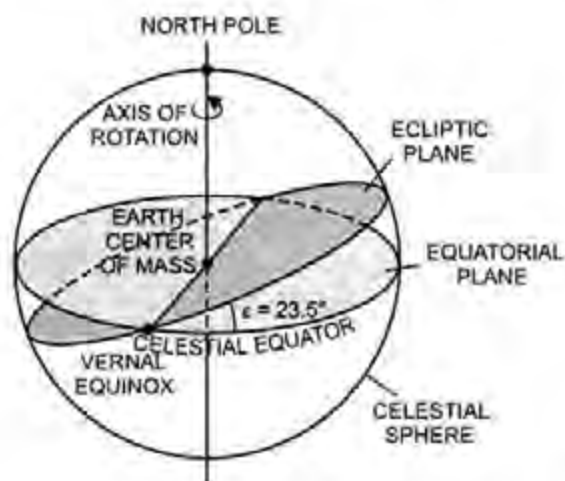
Fig. 2.1 Earth rotation, equatorial plane, and ecliptic plane

- the *diurnal rotation* (spin) of the earth about its polar axis. This rotational axis approximately coincides with the axis of maximum moment of inertia, and it passes through the earth's center of mass, cf. [2.5.2]. The *equatorial plane* is perpendicular to the axis of rotation,
- the *annual revolution* of the earth around the sun. Following Kepler's laws, the earth describes an ellipse with the sun at one of its focal points. Minor perturbations arise due to the gravitation of the moon and other planets. The plane of the earth's orbit is called the *ecliptic plane*; it has an obliquity of about 23.5° with respect to the equatorial plane.

By circumscribing the unit sphere around the center of the earth, simple geometric relations are obtained. The *celestial equator* and the *ecliptic* are defined by the intersections of the sphere with the corresponding planes. The *vernal equinox* (also first point of Aries) is the intersection of the ecliptic and the equator where the sun passes from the southern to the northern hemisphere.

Sidereal time is directly related to the rotation of the earth. *Local Apparent* (or true) *Sidereal Time* (LAST) refers to the observer's (local) meridian; it is equal to the hour angle of the (true) vernal equinox (Fig. 2.2), cf. [2.4.1]. The vernal equinox is affected by precession and nutation and experiences long and short-periodic variations, cf. [2.4.2]. If nutation is taken into account, we obtain *Local Mean Sidereal Time* (LMST), referring to the mean vernal equinox. For the Greenwich meridian the corresponding hour angles are called *Greenwich Apparent Sidereal Time* (GAST) and *Greenwich Mean Sidereal Time* (GMST). The astronomic longitude Λ is the angle between the meridian planes of the observer and Greenwich. It is given by, cf. [2.6.2]
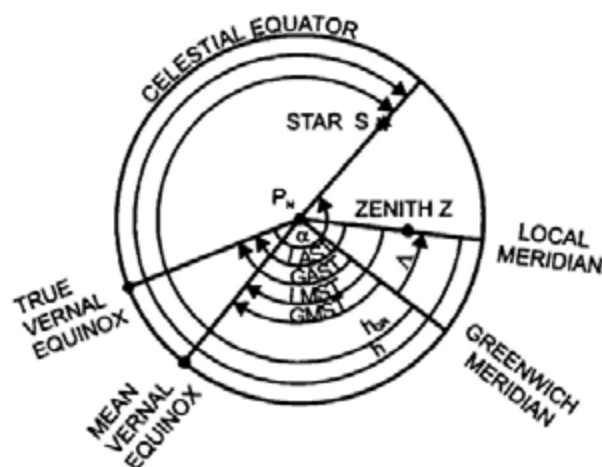
Fig. 2.2. Rectascension, sidereal time, hour angle, and longitude

$$\Lambda = \text{LAST} - \text{GAST} = \text{LMST} - \text{GMST} . \qquad (2.5)$$

LAST is determined from astronomical observations to fixed stars and extragalactic radio sources. The mean sidereal time scale is still affected by precession (long-periodic). The *mean sidereal day* is the fundamental unit; it corresponds to the time interval of two consecutive transits of the mean vernal equinox through the meridian.

For practical reasons, *solar time* is used in everyday life. It is related to the apparent diurnal motion of the sun about the earth. Since this revolution is not uniform, a "mean" sun is introduced which moves with constant velocity along the equator and coincides with the true sun at the vernal equinox. *Mean solar time* is equal to the hour angle of the mean sun plus 12 hours. If referred to the Greenwich mean astronomical meridian, cf. [2.5.1], it is termed *Universal Time* (UT). Its fundamental unit is the *mean solar day*, being the interval between two transits of the fictitious sun through the meridian.

The conversion of Universal Time to Greenwich Mean Sidereal Time is rigorously possible and is given by a series development with time defined by the International Astronomical Union (MORITZ and MUELLER 1987). Since the orbital motion of the earth is about 1° per day $(360°/365\,\text{d})$, the year has one day more in sidereal days than in solar days. We have the following approximation:

$$1 \text{ mean sidereal day} = 1 \text{ mean solar day} - 3\,\text{m}\,55.90\,\text{s} = 86164.10\,\text{s} . \quad (2.6)$$

The earth's rotation rate is 15.041 07"/1 s, and its angular velocity is

$$\omega = 2\pi/86\,164.10\,\text{s} = 7.292\,115 \times 10^{-5}\,\text{rad}\,\text{s}^{-1} . \qquad (2.7)$$

Universal time is obtained from a network of stations operating within the frame of the International Earth Rotation Service, cf. [2.3]. The observed local time UT0 refers to the instantaneous rotation axis, which is affected by polar motion, cf. [2.5.2]. In order to compare the results of different stations, reductions to a *Conventional Terrestrial Pole* are applied. The reduction in astronomic longitude $\Delta\Lambda_P$ corresponds to a change in time, cf. [5.3.3]. It transforms UT0 to UT1, which refers to the conventional terrestrial system, cf. [2.5.3]:

$$\text{UT1} = \text{UT0} + \Delta\Lambda_p . \qquad (2.8)$$

The precision of UT1 is about 0.01 to 0.02 ms at a 1d resolution.

UT1, as well as Greenwich Mean Sidereal Time, still contains the variations of the earth's rotation with time, which are of secular, periodic, and irregular character, cf. [2.5.2]. An approximation to a uniform time scale can be achieved by modeling the seasonal variations of annual and semiannual type. With the corresponding reduction $\Delta\Lambda_S$, we obtain

$$UT2 = UT0 + \Delta\Lambda_p + \Delta\Lambda_s . \qquad (2.9)$$

A practical time scale, as needed in navigation for instance, has to provide a uniform unit of time and maintain a close relationship with UT1. This led to the introduction of the *Coordinated Universal Time* (UTC). Its time interval corresponds to atomic time TAI, cf. [2.2.1], and its epoch differs by not more than 0.9s from UT1. In order to keep the difference

$$|DUT1| = |UT1 - UTC| < 0.9s , \qquad (2.10)$$

"leap seconds" are introduced to UTC when necessary. UTC is provided by the BIPM Time Section and broadcasted by time signal stations, while DUT1 is calculated by the IERS, cf. [2.3].

Among the continuously broadcasting time stations are DCF77/Mainflingen (77.5 kHz), HBG/Prangins (75 kHz); MSF/Rugby (60 kHz) in Europe; WWV resp. WWVB/Ft. Collins, Colorado (2500 to 20000 kHz resp. 60 kHz); and WWVH/Kauai, Hawaii (2500 to 15000 kHz).

2.3   International Earth Rotation Service

The International Earth Rotation Service (IERS) is in charge of providing and maintaining conventional celestial and terrestrial reference frames. These frames are a realization of the reference systems recommended by the International Astronomical Union (IAU) and the International Union of Geodesy and Geophysics (IUGG). IERS is also responsible for the determination of the orientation parameters of the earth as functions of time, which relate the two frames to each other (SEIDELMANN 1992, REIGBER and FEISSEL 1997).

Established by the IAU and IUGG, the IERS has operated since January 1, 1988. It collects, analyzes, and models observations of a global network of astronomic and geodetic stations (about 300 sites in 1996), operating either permanently or for a certain time span. Observation techniques include Very Long Baseline Interferometry (VLBI), Lunar Laser Ranging (LLR), Global Positioning System (GPS), Satellite Laser Ranging (SLR), and DORIS (Doppler Orbit determination and Radio positioning Integrated on Satellite), cf. [5.2] to [5.3].

The different types of observations are evaluated at the respective IERS coordinating centers and then combined by an adjustment at the IERS Central Bureau. The results include the positions (coordinates) of both the extragalactic radio sources and the terrestrial stations, the earth orientation parameters (EOP), and other information. With respect to the EOP, VLBI provides information about precession, nutation, polar motion, and UT1, cf. [2.4.2], [2.2.2]. Satellite techniques contribute to the daily interpolation of UT and to the determination of polar motion, cf. [2.5.2]. The results are disseminated through bulletins, annual reports, and technical notes. The combined solutions have an accuracy of ±0.0003″ for EOP and ±0.01 m for the positions of the terrestrial stations, cf. [2.4], [2.5]. The evaluation of the observations is based on the IERS Conventions, which are consistent with the IAU and IUGG/IAG recommendations for reference systems (McCARTHY 1996), cf. [2.4.2], [4.3].

Among the early international agreements on positioning and time were the introduction of the Greenwich zero meridian and Universal Time (1884). The first fundamental catalogue of selected stars was published in the 1880's, which started a series of star catalogues providing the positions of fixed stars.

International activities of monitoring the earth's rotation date back to 1899, when the *International Latitude Service* (ILS) started to determine polar motion through latitude observations at five observatories located around the globe on the 39°08′ northern parallel. After extension to the *International Polar Motion Service* (IPMS), and in cooperation with the *Bureau International de l'Heure* (BIH) established in 1912, about 50 astronomical observatories contributed to the determination of polar motion and time. An accuracy of ±0.02″ resp. ±1 ms was reached for mean values over 5 days. IPMS and the earth rotation section of BIH have been replaced by IERS, while the BIH activities on time are continued at the BIPM, cf. [2.2.1].

2.4   Celestial Reference System

An inertial system is needed in order to describe the motions of the earth and other celestial bodies in space, including those of artificial satellites. Such a system is characterized by Newton's laws of motion; it is either at rest or in the state of a uniform rectilinear motion without rotation. A space-fixed system (celestial reference system) represents an approximation to an inertial system and can be defined by appropriate conventions: Conventional Inertial System (CIS). The coordinate frame for such a system is provided by spherical astronomy [2.4.1]. The spatial orientation of this frame varies with time, and therefore, modeling of the variations is required [2.4.2]. The International Celestial Reference Frame represents the realization of the celestial reference system [2.4.3], KOVALEVSKY et al. (1989), SEIDELMANN (1992).

2.4.1   Equatorial System of Spherical Astronomy

The coordinates of the celestial reference-system are defined by the *equatorial system* of spherical astronomy (MUELLER 1969, EICHHORN 1974). We introduce a three-dimensional Cartesian coordinate system with the origin at the center of mass of the earth (geocenter). The Z-axis coincides with the rotational axis of the earth. The X and Y-axes span the equatorial plane, with the X-axis pointing to the vernal equinox and the Y-axis forming a right-handed system (Fig. 2.3), cf. [2.2.2].
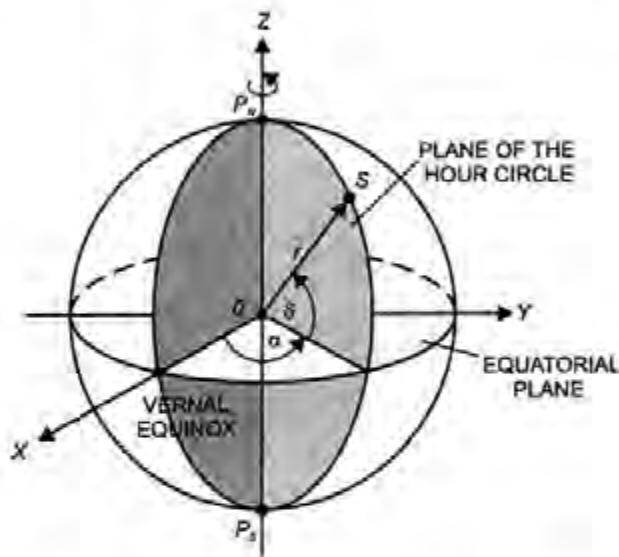
Fig. 2.3. Astronomic equatorial system

In the sequel, we shall also shift the origin of this system to the position of an observer on the earth (topocenter) or to the barycenter of the solar system. The directions to celestial bodies then vary with different definitions of the origin (parallaxes), cf. [5.3.3]. Since the earth's radius is negligibly small compared to the distances to stars and extragalactic radio sources, no distinction is necessary between a topocentric and a geocentric system.

We circumscribe the unit sphere (celestial sphere) about the earth. The rotational axis meets the sphere at the celestial north and south poles P_N and P_S. The great circles perpendicular to the celestial equator, which contain the celestial poles, are called *hour circles*, and the small circles parallel to the equator are termed *celestial parallels*.

The *right ascension* α is the angle measured in the plane of the equator between the planes of the hour circles passing through the vernal equinox and the celestial body S; it is reckoned from the vernal equinox anticlockwise. The

declination δ is the angle measured in the plane of the hour circle between the equatorial plane and the line OS (positive from the equator to P_N and negative to P_S).

The *position* of a *celestial body* S can be described either by the Cartesian coordinates X,Y,Z, or by the spherical coordinates α,δ,r (r = distance from the origin O). We have the transformation

$$\mathbf{r} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = r \begin{pmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{pmatrix}. \tag{2.11}$$

In geodesy, only directions are important for stars and extragalactic sources. With $r = 1$, α and δ describe the position of S on the unit sphere. They can also be expressed by the lengths of the corresponding arcs on the equator and the hour circle.
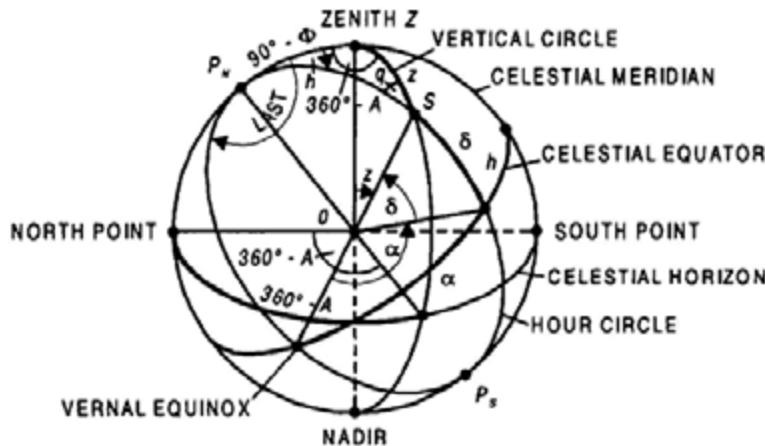
Fig. 2.4. Astronomic equatorial and horizon system

We introduce the local *meridian plane* of the observer, spanned by the local vertical (direction of the plumb line) and the rotational axis, after a parallel shift from the geocenter to the topocenter. The *zenithal point* Z is the intersection of the vertical with the unit sphere, and the *celestial meridian* is the great circle through Z and the poles (Fig. 2.4). The *hour angle h* is measured in the equatorial plane between the celestial meridian through Z and the hour circle of S, reckoned from the upper meridian toward west. Because of the earth's rotation, the hour angle system (h,δ) depends on time. The h,δ-system is rotated, with respect to the α,δ-system, about the polar axis by the angle of sidereal time LAST, cf. [2.2.2]. We have the relation (Fig. 2.2)

$$LAST = h + \alpha ,$$ (2.12)

which is used with time determination, cf. [5.3.2].

2.4.2 Precession and Nutation

The earth's axis of rotation, which has been introduced as the Z-axis, changes its spatial orientation with time. As a consequence, the position (α, δ) of a celestial body varies, with a superposition of long and short-periodic effects (MORITZ and MUELLER 1987, SEIDELMANN 1992, DICKEY 1995).

Fig. 2.5. Precession and nutation

The *lunisolar precession* is a long-periodic effect caused by the gravitation of the moon and the sun on the equatorial bulge of the earth. This creates a force couple (torque) which tends to turn the equatorial plane into the plane of the ecliptic (Fig. 2.5). In combination with the moment of the earth's rotation, the earth's axis describes a gyration of a cone with a generating angle of 23.5° (corresponding to the obliquity of the ecliptic ε), about the northern pole of the ecliptic E_N. The vernal equinox moves clockwise along the ecliptic at a rate of 50.3''/year, making a complete revolution in about 25800 years. The gravitation of the planets causes a slow dislocation of the earth's orbit and thereby an additional migration of the vernal equinox along the equator and a change in ε: *planetary precession*. The sum of the lunisolar and the planetary precession is termed *general precession*.

The precession is superimposed by short-periodic effects known as *nutation*, which has periods between 5 days and 18.6 years. These periods are mainly due

to the time variations of the inclination of the moon's orbit with respect to the ecliptic (appr. 5°). Other components have semiannual and semimonthly periods and stem from the oscillations of the sun and moon between the earth's northern and southern hemisphere.

Precession and nutation can be modeled as a function of time using the ephemerides of the moon, the sun, and the planets. The IAU (1976) theory of *precession* provides three time-dependent Eulerian rotation angles for reducing the positions of celestial bodies to a common reference. For the reference epoch J 2000.0 (Julian epoch January 1, 2000, 12^h TDB, cf. [2.2.1]), we have the fundamental constants "general precession in longitude at the ecliptic" (5029.0965"/century) and "obliquity of the ecliptic" (23°26'21.412").

The IAU (1980) theory of *nutation* describes this effect by a rotation about the cone of precession. The deviation of the true pole from the mean pole is modeled by two time-dependent parameters. Hereby, the earth is regarded as an elliptical, rotating, elastic, and ocean-free body with solid inner and liquid outer cores (WAHR 1981, SEIDELMANN 1992). For the epoch J 2000.0 the constant of nutation is 9.2025".

The IAU models for precession and nutation define the reference pole for the international celestial reference frame (Celestial Ephemeris Pole CEP). CEP is free of diurnal or quasidiurnal nutation terms (amplitudes < 0.001") with respect to the space- or earth-fixed coordinate systems. It is also referred to as the pole of the instantaneous equatorial system, cf. [2.4.3].

The IAU models for precession and nutation provide a precision of ±0.001" at 5 to 7 days resolution. An improved theory has been developed at the IERS based on recent VLBI and LLR data. Larger offsets (< 0.02") of the celestial pole from CEP have been found; these are published regularly by IERS (McCARTHY 1996). GPS also contributes to the determination of the short-periodic nutation terms (ROTHACHER et al. 1999).

According to an IAU recommendation, the IAU (1976/1980) models for precession and nutation shall be replaced by January 1, 2003. The new model is the IAU 2000A (precision ±0.0002"), as published in the IERS conventions 2000. CEP will then be substituted by the Celestial Intermediate Pole (CIP), as defined by the model for periods greater than two days together with additional time-dependent corrections provided by IERS.

The instantaneous position of a celestial body, cf. [2.4.1], is called *true position* at the epoch *t*. By accounting for nutation, we obtain the *mean position* at epoch *t*, which refers to the mean celestial equator and the mean vernal equinox, cf. [2.2.2]. If precession is also taken into account, we get the mean position at the reference epoch J 2000.0.

2.4.3   International Celestial Reference Frame

The International Celestial Reference System (ICRS), as recommended by IAU, is based on the general theory of relativity, with the time coordinate defined by the international atomic time (ARIAS et al. 1995, MA and FEISSEL 1997). ICRS approximates a space-fixed conventional inertial-system (CIS) with the origin at the barycenter of the solar system. It is assumed that no global rotation of the system exists. This implies that the defining sources are either free from proper motion (component of spatial motion tangent to the celestial sphere) or that this motion can be modeled. The coordinate axes are defined by the celestial reference pole and the vernal equinox as provided by the IAU models for precession and nutation. They are realized through mean directions to extraterrestrial fiducial objects: stellar or radio source CIS (MUELLER 1988), cf. [2.4.2].

The *stellar system* is based on the stars of the Fundamental Catalogue FK5 (FRICKE et al. 1988). It provides the mean positions (α, δ) and the proper motions (generally < 1"/year) of 1535 fundamental stars for the epoch J 2000.0, with precisions of ±0.01...0.03" and ±0.05"/century respectively. A supplement to FK5 contains additional stars up to an apparent magnitude of 9.5. The mean equator and the mean vernal equinox for J 2000.0 are realized by the FK5 catalogue, with an accuracy of ±0.05". Due to refraction uncertainties, earth-based astrometry can hardly improve this accuracy.

Astronomic space missions have significantly improved the realization of a stellar CIS. The HIPPARCOS astrometry satellite (ESA, 1989 – 1993) was used to construct a network by measuring large angles between about 100000 stars (up to an apparent magnitude of 9) covering the entire sky. The reference frame thus established provides an accuracy of ±0.001" and ±0.0005"/year for proper motion (HIPPARCOS 1995, KOVALEVSKY et al. 1997). From improved FK5 data and HIPPARCOS results, an FK6 catalogue has been developed for a small number of stars (340 "astrometrically excellent"), resulting in an improvement of proper motion as compared to the HIPPARCOS catalogue (WIELEN et al. 1999). Future astrometric space missions will employ optical interferometry and thus increase the positional accuracy to ±0.00001" (BROSCHE and DICK 1996).

The *radio source system* is based on extragalactic radio sources (quasars and other compact sources). It was adopted as ICRS by the IAU in 1997 and has superseded the previous stellar system (FK5) since 1998. Due to the large distances (> 1.5 billion light years), these sources do not show a measurable proper motion. The system is realized through the International Celestial Reference Frame (ICRF), established and maintained by IERS (McCARTHY 1996, MA et al. 1998). ICRF contains the coordinates (equatorial system, epoch J 2000.0) of more than 600 objects. About 200 of them are well observed
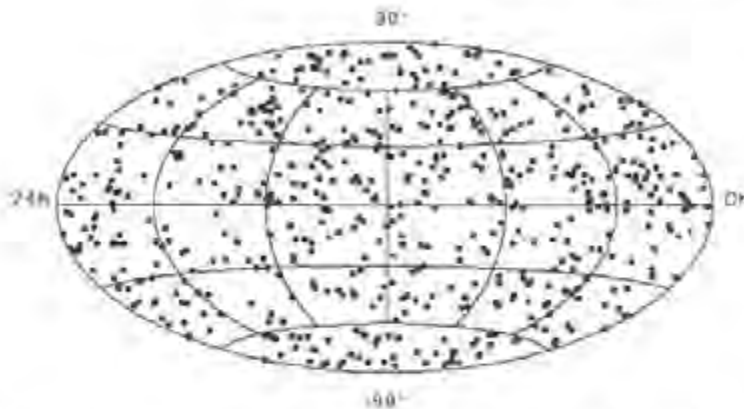
Fig. 2.6. International Celestial Reference Frame (ITRF), from IERS (1995): Missions and goals for 2000

"defining sources," and 100 more are used for densification and connection to the stellar-fixed reference system (Fig. 2.6). The southern sky is not as well covered, as the telescopes are concentrated in the northern hemisphere. The coordinates of the radio sources are determined by radio astronomy, with a precision of better than $\pm0.001''$ on the average and $\pm0.0003''$ for the most precisely observed objects (MA and FEISSEL 1997, BROSCHE und SCHUH 1999).

The link between the stellar and the radio source CIS is given with an accuracy of $\pm0.05...0.1''$, which corresponds to the uncertainty of FK5. This connection will be improved by the results of the astrometric space missions (optical signals from a limited number of radio sources) to $\pm0.001''$ or better for the epoch of observation.

2.5   Terrestrial Reference System

An *earth-fixed* reference system is introduced for positioning and navigation on and close to the earth's surface and for describing the earth's gravity field as well as other physical parameters. It is defined by a three-dimensional geocentric coordinate system [2.5.1]. The orientation of this system changes with time and with respect to the solid earth's body as well as to the celestial reference system [2.5.2]. The system is realized by the IERS International Terrestrial Reference Frame, which includes its relation to the International Celestial Reference Frame [2.5.3].

2.5.1   Global Earth-Fixed Geocentric System

An earth-fixed (i.e., rotating with the earth) system of spatial Cartesian-coordinates X,Y,Z is used as the fundamental terrestrial coordinate system (Fig. 2.7). Its origin is at the earth's center of mass (geocenter), being defined for the whole earth including hydrosphere and atmosphere. The Z-axis is directed towards a conventional "mean" terrestrial (north) pole. The "mean" equatorial plane is perpendicular to it and contains the X and Y-axes. A "mean" rotational axis and equatorial plane has to be introduced because the rotation of the earth changes with respect to the earth's body over time, cf. [2.5.3]. The XZ-plane is generated by the conventional "mean" meridian plane of Greenwich, which is spanned by the mean axis of rotation and the Greenwich zero meridian, to which Universal Time refers, cf. [2.2.2]. The Z and X-axes are realized indirectly through the coordinates of terrestrial "fiducial" stations, cf. [2.5.3]. The Y-axis is directed so as to obtain a right-handed system.
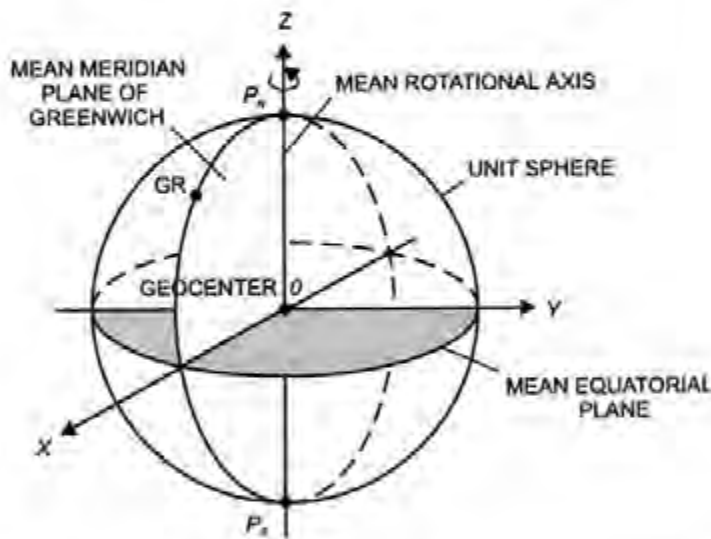
Fig. 2.7. Earth-fixed geocentric Cartesian system

The instantaneous axis of rotation is the common starting point for defining the Z-axes of the space-fixed and the earth-fixed reference systems. By referring to a reference epoch, the dependence on time is modeled, cf.[2.4.2], [2.5.2]. The directions of the X-axes of both systems differ by the angle of Greenwich mean sidereal time GMST, cf. [2.2.2].

In order to describe analytically certain physical properties of the earth (gravity field, magnetic field, topography, etc.), *spherical coordinates* r,ϑ,λ are employed. Here r = radial distance from the geocenter, ϑ = polar distance (colatitude), and λ = geocentric longitude. Instead of ϑ, the geocentric latitude
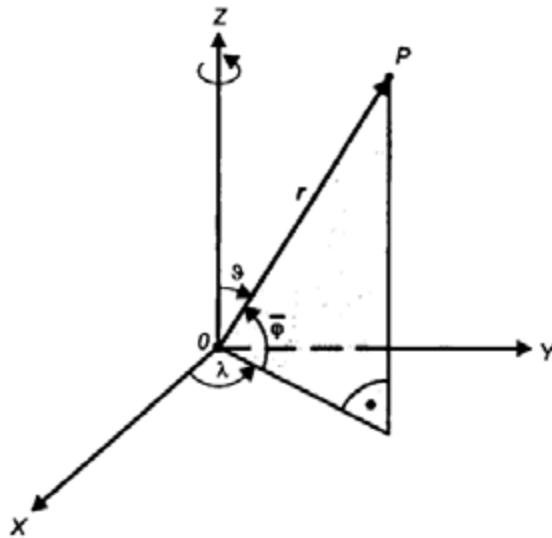
Fig. 2.8. Cartesian and spherical coordinates

$$\bar{\varphi} = 90° - \vartheta \qquad (2.13)$$

can be used (Fig. 2.8). The position of the point P is then given by the positional vector

$$\mathbf{r} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = r \begin{pmatrix} \sin\vartheta\cos\lambda \\ \sin\vartheta\sin\lambda \\ \cos\vartheta \end{pmatrix}. \qquad (2.14)$$

2.5.2   Polar Motion, Length of Day, Geocenter Variations

The rotation of the earth can be described by a vector directed to the north pole of the instantaneous axis of rotation and by the angular velocity ω, see (2.7).

Direction and magnitude of the rotational vector *change with time* due to astronomical and geophysical processes. These processes include variations of the lunar and solar gravitation and mass redistributions in the atmosphere, the hydrosphere, the solid earth, and the liquid core. The changes are secular, periodic or quasiperiodic, and irregular in nature (LAMBECK 1980, MORITZ and MUELLER 1987, DICKEY 1995), cf. [8.3.1].

Polar motion (or wobble) is the motion of the rotation axis relative to the earth's crust as viewed from the earth-fixed reference system. It directly affects

the coordinates of stations on the earth's surface and the gravity vector. Polar motion consists of several components:

- A free oscillation with a period of about 435 days (*Chandler period*), with an amplitude of 0.1″ to 0.2″, in a counterclockwise sense as viewed from the north pole. The Chandler wobble is due to the fact that the spin axis of the earth does not coincide exactly with a principal axis of inertia.

  For a rigid earth, this would lead to a gyration of the rotational axis about the principal axis of inertia with a period of $A/(C-A)=305$ days (Euler period). Here C is the earth's polar moment of inertia, and $A = B$ is the mean equatorial moment (rotational symmetry assumed). The difference between the Chandler and the Euler period results from the non-rigidity of the earth. This consideration neglects the fact that there is a small deviation between the axis of rotation and the axis of angular momentum, which is invariable in space. However, the deviation is less than 0.001″ with periods < 1 d.

- The Chandler wobble is superposed by an *annual oscillation* forced by seasonal displacements of air and water masses. It proceeds in the same direction as the Chandler wobble with amplitudes of 0.05″ to 0.1″.
- A *secular motion* of the pole has been observed for more than 100 years. The motion consists of an irregular drift of about 0.003″/year in the direction of the 80° W meridian. Secular motion is mainly due to the melting of the polar ice and to large-scale tectonic movements; it attains large amounts over geological epochs: *polar wander*.
- More *irregular variations* occur at time scales from a few days to years with amplitudes up to 0.02″. They originate primarily from mass redistributions within the atmosphere, but variations due to ocean volume changes, ground water variations, and earthquakes also occur.
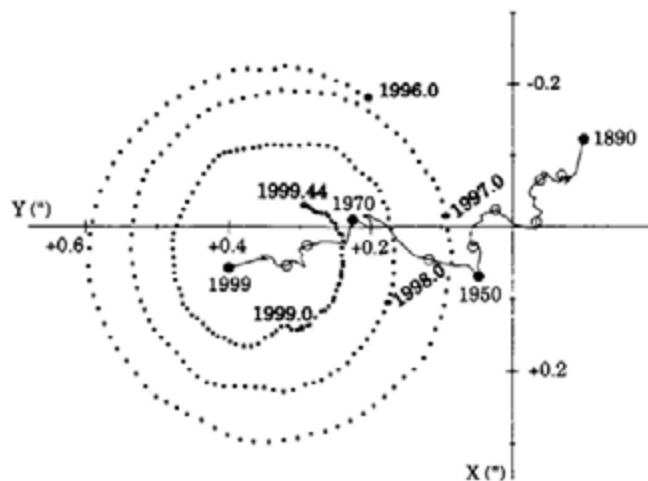
Fig. 2.9. Polar motion 1996 – 1999, and mean pole displacement 1890 – 1999, from IERS Annual Report 1998

The superposition of these components results in a slightly perturbed spiral like curve of the instantaneous pole with a slowly advancing mean position (Fig. 2.9). Over one year, the deviations from the mean position remain $< 0.3''$, corresponding to 9 m on the earth's surface.

The reference for describing the actual position of the pole with respect to the solid earth is provided by the *IERS reference pole*. It agrees within $\pm0.03''$ with the *Conventional International Origin* (CIO), which was defined by the mean position of the north pole as determined between 1900.0 and 1906.0. The position of the instantaneous pole (Celestial Ephemeris Pole, cf. [2.4.2]) with respect to the reference pole is given by the rectangular coordinates x_P, y_P, which are defined in the plane tangential to the pole. The x_P-axis is in the direction of the Greenwich mean meridian (consistent with the previous BIH zero meridian), and the y_P-axis is directed along the 90°W meridian. These plane coordinates are usually expressed as spherical distances (in units of arcsec) on the unit sphere.

The *angular velocity* ω of the earth's rotation, as monitored from the earth, changes with time. Relative changes may reach several 10^{-8}, which corresponds to several ms for one day. The variations are generally described by the excess revolution time with respect to 86 400 s and then called *Length Of Day* (LOD). They are derived by comparing astronomical time determinations, which deliver Universal Time UT1, with the uniform time scales TAI or UTC, cf. [2.2.2].
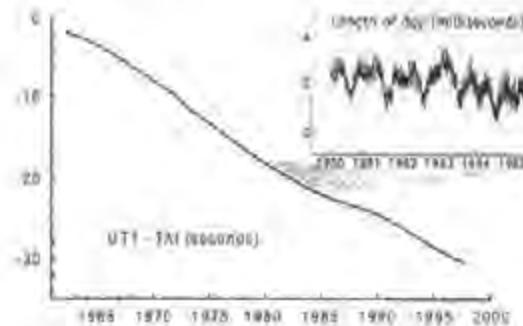
Fig. 2.10. Difference between atomic time scale TAI and Universal time UT1 (1962-1998) and length of day (LOD 1979-1987), from IERS Inform. 1998

The following components of LOD variations have been observed (Fig. 2.10):

- A *secular decrease* in the angular velocity of the earth's rotation is caused mainly by tidal friction. It lengthens the day by about 2 ms/century (STEPHENSON and MORRISON 1994).

- Fluctuations over *decades* are due to motions in the earth's liquid core and to slow climatic variations.
- The *tides* of the solid earth and the oceans produce variations of about 1 ms with long (annually) and short (monthly and less) periodic parts.
- *Seasonal* effects are explained by atmospheric excitation, with contributions from water and ice budget variations.
- More *irregular* oscillations stem from different sources, such as terrestrial mass displacements (earthquakes), solar activity, and atmospheric events, e.g., El Niño.

While the effect of polar motion on observations is dependent on location, LOD changes act uniformly on all points. The pole coordinates and LOD, as well as ω, are provided as *Earth Orientation Parameters* (EOP) by the IERS with daily resolution and accuracy of $\pm 0.0003''$ resp. ± 0.02 ms or better (REIGBER and FEISSEL 1997).

The position of the *geocenter* (origin of the terrestrial reference system) changes slightly in time with respect to the monitoring observatories. Annual and semiannual variations have been found, with amplitudes of several mm/year, from the analysis of satellite orbits. The variations are caused primarily by mass redistributions in the atmosphere and the oceans and by continental water variations. Through the coordinates of the ITRF stations, cf. [2.5.3], the geocenter is given with an accuracy of a few mm (DONG et al. 1997, RAY 1999).

2.5.3   International Terrestrial Reference Frame

The International Terrestrial Reference System is realized by the IERS through a global set of space geodetic observing sites. The geocentric Cartesian coordinates and velocities of the observing sites comprise the International Terrestrial Reference Frame (ITRF). The stations participating in the ITRF carry out observations either continuously or at certain time intervals (Fig. 2.11). Observations are made on twelve of the larger tectonic plates, which permits the derivation of station velocities related to plate tectonics, cf. [8.2.3].

Annual realizations of the ITRF are published by the IERS. The ITRF97 is comprised of the geocentric positions (X, Y, Z) for more than 550 stations at about 320 sites and corresponding site velocities (BOUCHER et al. 1999). The accuracy of the results depends on the observation techniques and is maximum for VLBI, SLR, and GPS observations ($\pm 0.5...2$ cm and $\pm 1...3$ mm/year resp.). Several time variable effects are taken into account, including displacements due to the solid earth tides, ocean and atmospheric loading effects, and postglacial rebound, cf. [8.2.2]. The ITRF solutions satisfy the condition of no residual net-rotation relative to the plate tectonics model NNR-NUVEL1A; vertical movements are not allowed at all, cf. [8.2.3]. The orientation of the
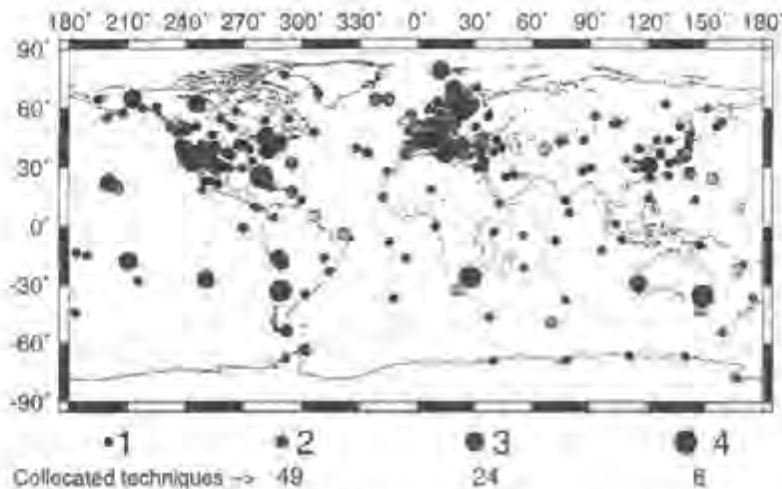
Fig. 2.11. International Terrestrial Reference Frame (ITRF) sites 1997, from BOUCHER et al. (1999)

ITRF is given with respect to the IERS reference pole and reference meridian, cf. [2.5.2]. The actual (time t) position vector **r** of a point on the earth's surface is derived from its position at the reference epoch (t_0) by

$$r(t) = r_0 + \dot{r}_0 (t - t_0),$$ (2.15)

where r_0 and $\dot{r}_0$ are the position and velocity respectively at t_0.

The relation between the celestial (ICRS) and the terrestrial (ITRS) reference systems is given by spatial rotations, which depend on the earth rotation parameters introduced in [2.2.2], [2.4.2], [2.5.2], SEEBER (1993), McCARTHY (1996), RICHTER, BU. (1995). The complete transformation from the celestial to the terrestrial system reads as

$$r(ITRS) = R_2(-x_p)R_1(-y_p)R_3(GAST)N(t)P(t)r(ICRS).$$ (2.16)

The position vector as given in the ICRS is first transformed by the precession matrix $P(t)$ from the reference epoch t_0 (J 2000.0) to the observation epoch t. The nutation matrix $N(t)$ then transforms from the mean to the instantaneous true equator and vernal equinox. The Eulerian angles in these two rotation matrices are given in the models for precession and nutation, cf. [2.4.2]. The apparent Greenwich sidereal time GAST, cf. [2.2.2], is used to rotate the system about the Z-axis:

$$\mathbf{R}_3\left(GAST\right)=\begin{pmatrix} \cos\left(GAST\right) & \sin\left(GAST\right) & 0 \\ -\sin\left(GAST\right) & \cos\left(GAST\right) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.17}$$

with GAST calculated from UT1. Finally, counterclockwise rotations about the X and Y-axes are computed as functions of the pole coordinates x_P and y_P (small angles), cf. [2.5.2] .

$$\mathbf{R}_1\left(-y_P\right)=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -y_P \\ 0 & y_P & 1 \end{pmatrix}, \quad \mathbf{R}_2\left(-x_P\right)=\begin{pmatrix} 1 & 0 & x_P \\ 0 & 1 & 0 \\ -x_P & 0 & 1 \end{pmatrix}. \tag{2.18}$$

Equations (2.17) and (2.18) provide the transformation from the instantaneous space-fixed system to the conventional terrestrial system.

2.6   Gravity Field Related Reference Systems

Most geodetic and astronomic observations on or close to the earth's surface refer to the earth's gravity field by orientation along the local vertical. Consequently, local gravity-field-related reference systems are introduced for the modeling of these observations. The orientation of the local systems with respect to the global reference system is given by astronomic latitude and longitude [2.6.1]. These orientation parameters are used for transformation from the local systems into the global system and back [2.6.2].

2.6.1   Orientation of the Local Vertical

The direction of the plumb line (local vertical) with respect to the global geocentric system is given by two angles (Fig. 2.12). The *astronomic (geographic) latitude* Φ is the angle measured in the plane of the meridian between the equatorial plane and the local vertical through the point P. It is reckoned positive from the equator northward and negative to the south. The angle measured in the equatorial plane between the Greenwich meridian plane and the plane of the meridian passing through P is the *astronomic (geographic) longitude* Λ; it is reckoned positive toward the east. The *gravity potential W* locates P in the system of level surfaces $W = $ const., cf. [3.2.1]. The *local astronomic meridian plane* is spanned by the local vertical at P and a line parallel to the rotational axis, cf. [2.4.1].
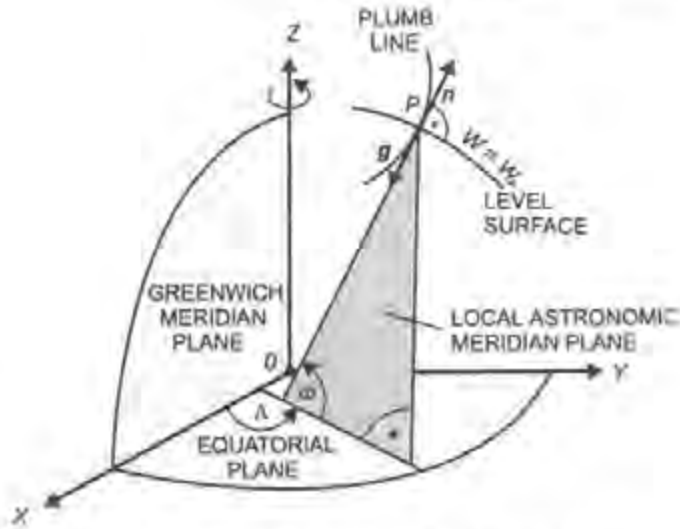
Fig. 2.12. Astronomic latitude and longitude

We introduce the outer surface normal **n** (unit vector), which is normal to the level surface $W = W_p$ and passes through P. It is directed to the zenith, which is opposite of the direction of the gravity vector **g**. From Fig. 2.12, we see that

$$\mathbf{n} = -\frac{\mathbf{g}}{g} = \begin{pmatrix} \cos\Phi\cos\Lambda \\ \cos\Phi\sin\Lambda \\ \sin\Phi \end{pmatrix}. \tag{2.19}$$

Latitude Φ and longitude Λ can be determined by the methods of geodetic astronomy, cf. [5.3]. Together with the potential W, they form a triple of three-dimensional coordinates defined in the gravity field, cf. [3.2.3].

2.6.2   Local Astronomic Systems

Geodetic and astronomic observations are tied to the direction of the plumb line at the point of observation and thereby to the earth's gravity field. An exception is distance measurements, which are independent of the reference system. Thus, these observations establish local gravity-field related systems: *Local astronomic systems* (Fig. 2.13). Their origin is at the point of observation P. The z-axis coincides with the local vertical and points toward the zenith. The x-axis (north) and the y-axis (east) span the horizontal plane, which is tangent to the level surface $W = W_P$. This x,y,z-system is left-handed.

Observable geometric quantities include astronomic azimuths, horizontal directions and angles, zenith angles, spatial distances, and leveled height differences.

The *astronomic azimuth A* is the angle measured in the horizontal plane between the astronomic meridian of P and the vertical plane spanned by the vertical through P and the target point P_i. It is positive as measured from the x-axis in a clockwise direction. *Horizontal directions* and *angles* may be regarded as azimuths lacking orientation, or as azimuth differences. The *zenith angle* (zenith distance) z is the angle measured in the vertical plane between the local vertical and the line joining P and P_i. It is positive as measured from the outer surface normal. The *spatial distance s* is the length of the straight line joining P and P_i. *Geometric leveling* also refers to the local vertical, providing a height difference with respect to $W = W_P$ over a very short distance. It may be regarded as the boundary case for trigonometric heighting, with a zenith angle of 90°. *Gravity measurements* and measurements of *gravity gradients* also refer to the local astronomic system.
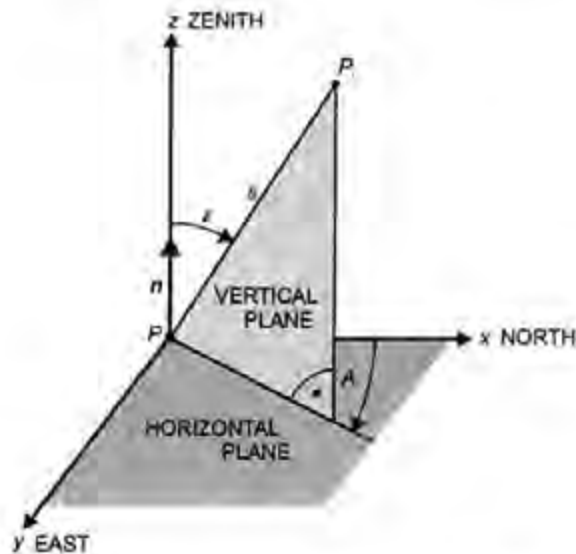
Fig. 2.13. Local astronomic system

According to Fig. 2.13, the position vector between P and P_i is given by

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = s \begin{pmatrix} \cos A \sin z \\ \sin A \sin z \\ \cos z \end{pmatrix}. \tag{2.20}$$

The local astronomic system is used for astronomic and geodetic applications.
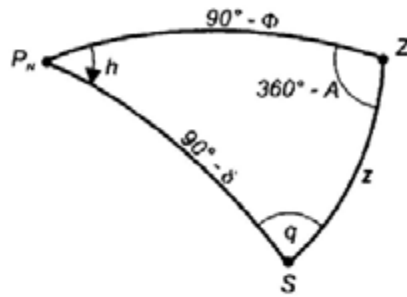
Fig. 2.14. Astronomic triangle

In *geodetic astronomy*, only direction measurements (zenith angles and azimuths) to celestial bodies are performed. The local system is called the *horizon system*, and the origin is named topocenter. The points of intersection of the plumb-line direction with the celestial sphere are known as the zenithal point Z and the nadir point Z'. The intersection of the horizontal plane with the celestial sphere is the celestial horizon. The azimuth in astronomy is usually reckoned from the south point and is considered positive westward to the north. The relation between the horizon system and the equatorial hour angle system, cf. [2.4.1], is given by the astronomic triangle (Fig. 2.14), see also Fig. 2.4. It is formed on the celestial sphere by the vertices P_N (north pole), Z (zenithal point), and S (celestial body). The triangle contains the compliments to declination ($90° - \delta$) and astronomic latitude ($90° - \Phi$), the hour angle h, the zenith angle z, the explement of the azimuth ($360° - A$), and the parallactic angle q. From spherical astronomy we obtain the transformations

$$\left.\begin{aligned}
\cos A \sin z &= \sin \delta \cos \Phi - \cos \delta \cos h \sin \Phi \\
\sin A \sin z &= -\cos \delta \sin h \\
\cos z &= \sin \delta \sin \Phi + \cos \delta \cos h \cos \Phi
\end{aligned}\right\}. \tag{2.21}$$

Here the azimuth A is reckoned in the geodetic sense, i.e., positive from the north.

The transition to the α,δ-system (α = right ascension) is given by the local apparent sidereal time LAST, see (2.12):

$$\alpha = \text{LAST} - h. \tag{2.22}$$

Astronomic longitude Λ is obtained by comparing LAST with the Greenwich sidereal time (2.5):

$$\Lambda = \text{LAST} - \text{GAST}. \tag{2.23}$$

Equations (2.21) to (2.23) are the fundamental equations for determining Φ, Λ and A from measurements of z and GAST at given α, δ, cf. [5.3.2]. Equation (2.21) also follows from (2.20) if we take (2.11) and (2.29) into account.

For *geodetic* applications, the observations carried out in the local astronomic systems have to be transformed into the global geocentric system for further use in establishing geodetic control networks.

Due to the non-parallelism of the plumb lines, the orientation of the local systems depends on position and thus changes rapidly from place to place. Computations in *one* individual system are therefore admissible only in very limited areas when applying formulas of plane geometry.

The plumb line direction can be referred to the global geocentric-system by means of the "orientation" parameters astronomic latitude Φ and longitude Λ (Fig. 2.15). After a parallel shift of the global system into the local one (Fig. 2.16), we transform the latter one to a right-handed system by applying the reflection matrix

$$S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.24}$$
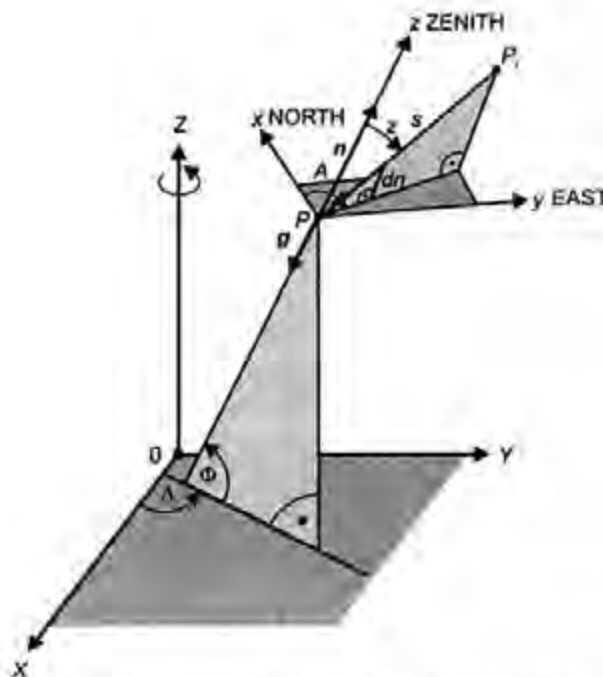
Fig. 2.15. Local astronomic and global geocentric system
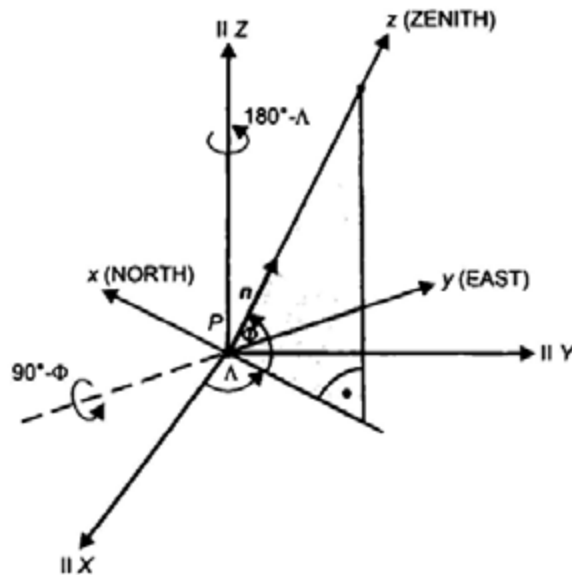
Fig. 2.16 Transformation between the local astronomic and the geocentric system

We then rotate the local system by $90° - \Phi$ around the (new) y-axis and by $180° - \Lambda$ around the z-axis with the rotation matrices

$$\mathbf{R}_2\left(90° - \Phi\right) = \begin{pmatrix} \sin\Phi & 0 & -\cos\Phi \\ 0 & 1 & 0 \\ \cos\Phi & 0 & \sin\Phi \end{pmatrix} \qquad \text{and}$$

$$\mathbf{R}_3\left(180° - \Lambda\right) = \begin{pmatrix} -\cos\Lambda & \sin\Lambda & 0 \\ -\sin\Lambda & -\cos\Lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (2.25)$$

Coordinates differences between P_i and P in the geocentric system are thus obtained by

$$\Delta\mathbf{X} = \mathbf{Ax}, \qquad (2.26)$$

with $\mathbf{x}$ given by (2.20) and

$$\Delta \mathbf{X} = \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix}. \tag{2.27}$$

The transformation matrix reads as

$$\mathbf{A} = \mathbf{R}_3 (180° - \Lambda) \mathbf{R}_2 (90° - \Phi) \mathbf{S}_2 =$$
$$\begin{pmatrix} -\sin\Phi\cos\Lambda & -\sin\Lambda & \cos\Phi\cos\Lambda \\ -\sin\Phi\sin\Lambda & \cos\Lambda & \cos\Phi\sin\Lambda \\ \cos\Phi & 0 & \sin\Phi \end{pmatrix}. \tag{2.28}$$

The inversion of (2.26) is performed easily considering that **A** is orthonormal:

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}.$$

We obtain

$$\mathbf{x} = \mathbf{A}^{-1} \Delta \mathbf{X}, \tag{2.29}$$

with

$$\mathbf{A}^{-1} = \begin{pmatrix} -\sin\Phi\cos\Lambda & -\sin\Phi\sin\Lambda & \cos\Phi \\ -\sin\Lambda & \cos\Lambda & 0 \\ \cos\Phi\cos\Lambda & \cos\Phi\sin\Lambda & \sin\Phi \end{pmatrix}. \tag{2.30}$$

Equations (2.27) to (2.30) are the basic equations for the evaluation of local geodetic measurements within the three-dimensional reference frame, cf. [6.2.1].

3   The Gravity Field of the Earth

The external gravity field plays a fundamental role in geodesy. This is because the figure of the earth has evolved under the influence of gravity, and most geodetic observations refer to gravity. Modeling of the observations thus requires knowledge of the gravity field. In addition, the analysis of the gravity field yields information on the structure of the earth's interior; in this way geodesy contributes to geophysics, cf.[8.2.4].

The fundamental quantities gravitation and gravity, together with their corresponding potentials, are introduced in [3.1], where the main properties of the gravity field are also described. The geometry of the gravity field is especially important for local applications [3.2], while the spherical harmonic expansion provides a powerful tool for a global gravity-field representation [3.3]. The geoid, as a physically defined reference surface for heights, is of basic interest in geosciences and engineering [3.4]. Gravity variations with time can be modeled in part (tidal effects), but to a large degree this work is still in the research stage [3.5].

The theory of the gravity field is extensively treated in geodetic and geophysical literature (e.g., HEISKANEN and VENING-MEINESZ 1958, HEISKANEN and MORITZ 1967, JEFFREYS 1970, PICK et al. 1973, GROTEN 1979).

3.1   Fundamentals of Gravity Field Theory

A body rotating with the earth experiences the gravitational force of the masses of the earth [3.1.1] to [3.1.3], and other celestial bodies as well as the centrifugal force due to the earth's rotation [3.1.4]. The resultant is the force of gravity [3.1.5]. In the case of artificial satellites, it is noted that a satellite does not rotate with the earth; hence, only gravitation acts on the satellite.

3.1.1   Gravitation, Gravitational Potential

According to Newton's law of gravitation (1687), two point masses m_1 and m_2 attract each other with the gravitational force (attractive force)

$$\mathbf{K} = -G \frac{m_1 m_2}{l^2} \frac{\mathbf{l}}{l},\qquad(3.1)$$

where G is the gravitational constant, cf. [2.1],

$$G = 6.673 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}, \tag{3.2}$$

and l is the distance between the masses. The vectors $\mathbf{K}$ and $\mathbf{l}$ point in opposing directions. The SI-unit of K is m kg s^{-2}.

By setting the mass at the attracted point P to unity, (3.1) transforms into the gravitational acceleration (henceforth also termed gravitation):

$$\mathbf{b} = -G \frac{m}{l^2} \frac{\mathbf{l}}{l}. \tag{3.3}$$
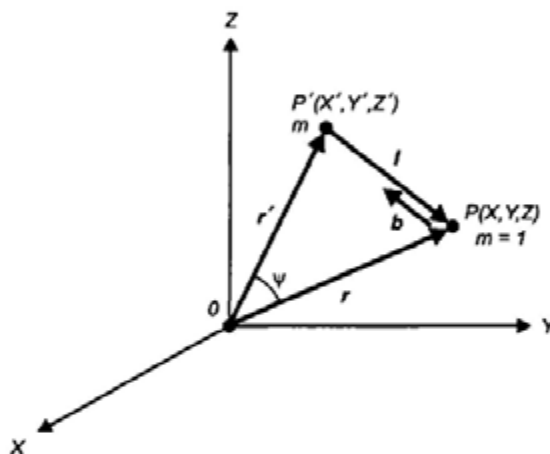
Fig. 3.1. Gravitation

$\mathbf{b}$ originates at P and is directed towards the source point P'. The vector $\mathbf{l}$ may be expressed by the position vectors $\mathbf{r}$ and $\mathbf{r}'$ (Fig. 3.1), e.g., in the global Cartesian X,Y,Z-system:

$$\mathbf{l} = \mathbf{r} - \mathbf{r}', \quad \mathbf{r}^T = (X, Y, Z), \text{ and } \mathbf{r}'^T = (X', Y', Z'), \tag{3.4}$$

with

$$l = |\mathbf{l}| = \sqrt{(X - X')^2 + (Y - Y')^2 + (Z - Z')^2} .$$

The unit of the acceleration b is ms^{-2}.

As gravitation depends only on the distance between the attracting mass and the attracted point, it does not depend on the choice of the coordinate system. While global applications require a

geocentric coordinate system, local coordinate systems are useful for problems of limited spatial extent.

The *earth* is composed of an infinite number of differential mass elements *dm*. The gravitation on the unit mass at P results from the integral over the individual contributions. Equation (3.3) correspondingly transforms to

$$\mathbf{b} = \mathbf{b}(\mathbf{r}) = -G\iiint\limits_{earth} \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} dm. \tag{3.5}$$

The mass element *dm* can be expressed by the volume *density* $\rho = \rho(\mathbf{r}')$ and the volume element *dv*:

$$dm = \rho \, dv, \tag{3.6}$$

where ρ is expressed in kg m^{-3}.

The representation of gravitational acceleration, the gravity field, and related computations are simplified if the scalar quantity "*potential*" is used instead of the vector quantity "*acceleration*." Because the gravitational field is invariant to rotations:

$$\text{curl}\,\mathbf{b} = \mathbf{0}, \tag{3.7}$$

the vector $\mathbf{b}$ may be represented as the gradient of a potential V (e.g., KELLOGG 1929, SIGL 1985):

$$\mathbf{b} = \text{grad}\,V. \tag{3.8}$$

For a *point mass m*, see (3.3), we have

$$V = \frac{GM}{l}, \text{ with } \lim_{r \to \infty} V = 0. \tag{3.9}$$

For the *earth*, see (3.5) and (3.6), we obtain

$$V = V(\mathbf{r}) = G\iiint\limits_{earth} \frac{dm}{l} = G\iiint\limits_{earth} \frac{\rho}{l} dv, \quad \lim_{r \to \infty} V = 0. \tag{3.10}$$

The potential at P indicates the work that must be done by gravitation in order to move the unit mass from infinity $(V = 0)$ to P. The unit of potential is m^{2}s^{-2}.

If the density function $\rho = \rho(r')$ was known for the earth, (3.5) resp. (3.10) would permit calculation of the gravitation as a function of position. In reality, more detailed density information is available for the upper layers of the earth only, while global models merely consider radial density changes, cf. [3.1.2], [8.1]. Consequently, gravity-field *observations* have to be used in order to model the exterior gravity field.

3.1.2   Gravitation of a Spherically Symmetric Earth

To a first approximation, the earth can be viewed as a sphere with a centrally symmetric density structure, i.e. composed of spherical shells with constant density, cf. [8.1]. We calculate the gravitation in the interior and exterior of this model using the system of spherical coordinates r, ϑ, λ introduced in (2.14). For this purpose, the orientation of the system is changed such that the ϑ-axis coincides with the line joining the coordinate origin O and the calculation point P (Fig. 3.2).
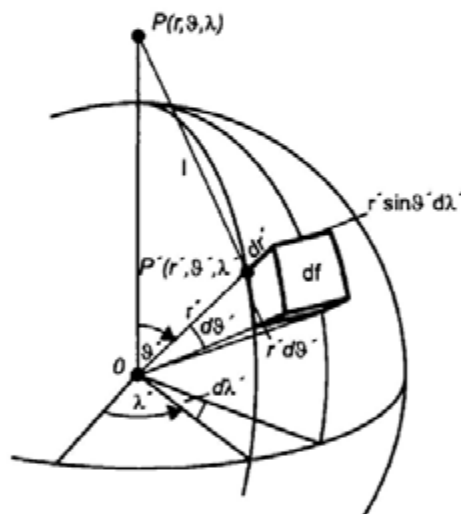
Fig. 3.2. Surface element of a spherical shell

The potential of a homogeneous *spherical shell* of radius r' with infinitesimal thickness dr' and density ρ is given in analogy to (3.10) by

$$V' = G\rho\,dr' \iint_f \frac{df}{l} .$$

(3.11)

Here, integration is over the surface of the shell f and

$$df = r'^2 \sin\vartheta'\,d\vartheta'\,d\lambda'$$

(3.12)

is the surface element. Inserting (3.12) into (3.11) gives

$$V' = G\rho r'^2 dr' \int\limits_{\lambda'=0}^{2\pi} \int\limits_{\vartheta'=0}^{\pi} \frac{\sin\vartheta'}{l} d\vartheta' d\lambda'. \tag{3.13}$$

A distinction is now made in the integration as to whether the attracted point P is exterior or interior to the spherical shell.

For an attracted point lying in the *exterior* $(r > r')$, the potential is given by integration of (3.13):

$$V'_e = 4\pi G\rho \frac{r'^2}{r} dr' = G\frac{dm'}{r}. \tag{3.14}$$

Here,

$$dm' = 4\pi\rho r'^2 dr' \tag{3.15}$$

represents the mass of the spherical shell. The potential of the *spherical earth* composed of concentric homogeneous shells is

$$V_e = G\iiint\limits_{earth} \frac{dm'}{r} = \frac{GM}{r}. \tag{3.16}$$

Hence, it is equal to the potential of the entire mass M of the earth concentrated at the center of mass. The gravitation follows from

$$b_e = -\frac{\partial V_e}{\partial r} = \frac{GM}{r^2}. \tag{3.17}$$

With $GM = 398.6 \times 10^{12}$ m^{3}s^{-2} and the radius of the earth $R = 6371$ km, the value of the potential at the surface of the earth $(r = R)$ amounts to $V = 6.26 \times 10^7$ m^{2}s^{-2}, and the gravitation is $b = 9.82$ ms^{-2}.

For a point in the *interior* $(r < r')$, we obtain from (3.13) for the potential of the spherical shell:

$$V'_i = 4\pi G\rho r' dr' = \frac{G\,dm'}{r'}. \tag{3.18}$$

Here, V' is constant; therefore, the gravitation is zero:

$$b_i' = -\frac{\partial V_i'}{\partial r} = 0. \tag{3.19}$$

The potential inside an *earth* constructed of *shells* includes the contribution of V_e' (3.14) due to the masses interior to the sphere $r = $ const. It is also comprised of V_i' (3.18), which is due to the spherical shell having thickness $R - r$:

$$V_i = \frac{4\pi G}{r} \int_0^r \rho r'^2 \, dr' + 4\pi G \int_r^R \rho r' \, dr'. \tag{3.20}$$

For a *homogeneous* earth ($\rho = $ const.) we have

$$V_i = \frac{4}{3}\pi G \rho r^2 + 2\pi G \rho \left(R^2 - r^2\right) = 2\pi G \rho \left(R^2 - \frac{r^2}{3}\right). \tag{3.21}$$

The gravitation of the earth composed of *shells* is

$$b_i = -\frac{\partial V_i}{\partial r} = G\frac{m'}{r^2} \tag{3.22}$$

with

$$m' = 4\pi \int_0^r \rho r'^2 \, dr' \tag{3.23}$$

according to (3.15), which represents the mass inside the sphere $r = $ const. The masses outside this sphere have no effect on the gravitation. For a *homogeneous* sphere, (3.22) and (3.23) transform to

$$b_i = \frac{4}{3}\pi G \rho r. \tag{3.24}$$

3.1.3   Properties of the Gravitational Potential

We now investigate the fundamental properties of the gravitational potential and its first and second derivatives.

Starting from the potential (3.10)

$$V = G \iiint_{earth} \frac{dm}{l}, \tag{3.25}$$

gravitation is given by the gradient (3.8) through the components, see (3.5),

$$\frac{\partial V}{\partial X} = V_X = -G \iiint_{earth} \frac{X - X'}{l^3} dm \text{ , etc.} \tag{3.26}$$

The second derivatives read as

$$\frac{\partial^2 V}{\partial X^2} = V_{XX} = -G \iiint_{earth} \frac{dm}{l^3} + 3G \iiint_{earth} \frac{(X - X')^2}{l^5} dm \text{ , etc.} \tag{3.27}$$

We now have to distinguish between the cases where the attracted point P lies exterior or interior to the earth's masses, cf. [3.1.2]. Here, we neglect the mass of the atmosphere (about 10^{-6} of the total mass) and the variations of gravitation with time (maximum relative effect about 10^{-7}). The earth's surface S then constitutes a boundary surface between the mass-free exterior space and the earth's interior.

If P lies exterior to the surface S, then $l > 0$ always. Then according to (3.25) - (3.27), the potential and its first and second derivatives are single-valued, finite and continuous functions, vanishing at infinity.

We now apply the Laplacian differential operator $\Delta = $ div grad to V. In the X,Y,Z-system, this reads as

$$\Delta V = V_{XX} + V_{YY} + V_{ZZ}. \tag{3.28}$$

When introducing (3.27) into (3.28), the first and second terms cancel each other. This leads to *Laplace's differential equation* of second order, which governs the exterior gravitational field:

$$\Delta V = 0. \tag{3.29}$$

Continuous functions having continuous first and second order derivatives and fulfilling (3.29) are called *harmonic functions*.

If the attracted body lies *inside* the body of the earth, then the case $l = 0$ is possible. This requires special attention because of the discontinuity of $1/l$.

To this end, we consider P enclosed by a sphere K (center at P_0, radius p),
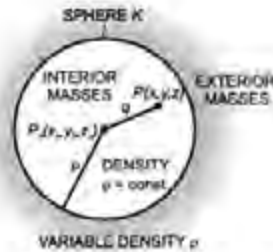
Fig. 3.3. Gravitational potential inside the earth

where p is chosen sufficiently small so that the density $\rho = $ const. inside K (Fig. 3.3). The potential at P is composed of the contributions from masses lying interior and exterior to K. From (3.10) and (3.21) and using

$$R = p, \quad r = q = \sqrt{(X-X_0)^2 + (Y-Y_0)^2 + (Z-Z_0)^2},$$

we find

$$V = G \iiint_{earth-K} \frac{dm}{l} + 2\pi G\rho\left(p^2 - \frac{q^2}{3}\right).$$

In the limits $p\to 0$ and $q\to 0$, agreement is obtained with the expression for the exterior potential (3.10). Differentiation yields

$$V_X = -G \iiint_{earth-K} \frac{X-X'}{l^3} dm - \frac{4}{3}\pi G\rho(X-X_0), \text{ etc.}$$

As $q\to 0$, we also have $X-X_0 = Y-Y_0 = Z-Z_0 \to 0$, so that once again we obtain agreement with the exterior case (3.26). The second derivatives are given by

$$V_{XX} = -G \iiint_{earth-K} \frac{1}{l^3} dm + 3G \iiint_{earth-K} \frac{(X-X')^2}{l^5} dm - \frac{4}{3}\pi G\rho, \text{ etc.}$$

For $q\to 0$, the last term does not vanish:

$$V_{XX} = -\frac{4}{3}\pi G\rho, \text{ etc.} \tag{3.30}$$

The gravitational potential and its first derivatives are thus single-valued, finite, and continuous in the interior as well. According to (3.30), the second derivatives exhibit discontinuities at abrupt changes in density. Inserting (3.30) into (3.28), we get *Poisson's differential equation*:

$$\Delta V = -4\pi G\rho.$$

(3.31)

Hence, V is not a harmonic function in the interior of the earth.

Finally, we mention *Gauss' integral formula*, which connects the normal derivatives $\partial V / \partial n_s$ on any boundary surface S (which in general is not an equipotential surface) and the second derivatives contained in the Laplace operator (3.28):

$$\iint_S \frac{\partial V}{\partial n_s} dS = \iiint_v \Delta V \, dv.$$

(3.32)
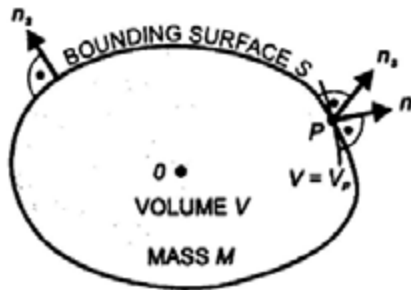
Fig. 3.4. Outer surface normal on the bounding surface and on the equipotential surface

Here, v is the volume of the body of surface S (Fig. 3.4). The left-side term may be interpreted as "gravitational flux" through S. As shown in potential theory, it is proportional to the total mass

$$M = \iiint_v dm = \iiint_v \rho(\mathbf{r}')dv$$

(3.33)

according to

$$\iint_S \frac{\partial V}{\partial n_s} dS = -4\pi GM.$$

(3.34)

Taking the limit at the source point P' in (3.34), equation (3.32) reduces to Poisson's differential equation (3.31) and to Laplace's differential equation for the exterior space ($\rho = 0$). Based on Gauss' formula, basic relationships can be established between observations in the gravity field and parameters describing the surface S, cf. [6.5.1].

3.1.4 Centrifugal Acceleration, Centrifugal Potential

The centrifugal force acts on an object of mass on the earth's surface. It arises as a result of the rotation of the earth about its axis. We assume here a rotation of constant angular velocity ω about the rotational axis, with the axis assumed fixed with the earth. The small effects of time variations of the rotational vector can be taken into account by reductions, cf. [2.5.2]. The centrifugal acceleration

$$z = (\omega \times r) \times \omega = \omega^2 p \qquad (3.35a)$$

acting on a unit mass is directed outward and is perpendicular to the spin axis (Fig. 3.5). With the geocentric latitude $\bar{\varphi}$, we have

$$p = r \cos \bar{\varphi}$$

and the magnitude

$$z = |z| = \omega^2 r \cos \bar{\varphi} . \qquad (3.35b)$$
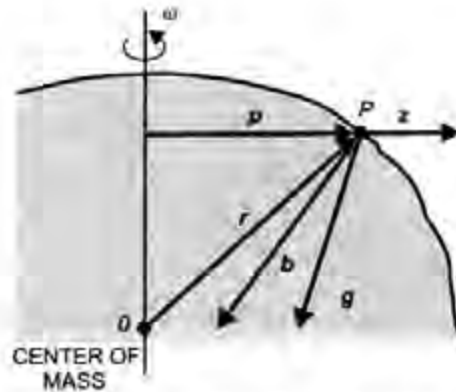
Fig. 3.5. Gravitation, centrifugal acceleration, and gravity

The angular velocity

$$\omega = 7.292\,115 \times 10^{-5} \,\mathrm{rad\,s^{-1}} \qquad (3.36)$$

is known with high accuracy from astronomy, cf. [2.2.2]. Consequently, $z = |\mathbf{z}|$ can be calculated if the position of P is known.

As the Z-axis of the earth-fixed X,Y,Z-system coincides with the axis of rotation, cf. [2.5.1], we have

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

With

$$\mathbf{z} = \operatorname{grad} Z \;, \tag{3.37}$$

we introduce the *centrifugal potential*

$$Z = Z(p) = \frac{\omega^2}{2} p^2 \;, \quad \lim_{p \to 0} Z = 0 \;. \tag{3.38}$$

Remark: Here, the symbols z and Z are used for the centrifugal acceleration and potential respectively. They were introduced earlier for local and global coordinates and will be employed again as such in later sections.

Differentiating twice and applying the Laplacian operator yields

$$\Delta Z = 2\omega^2 \;. \tag{3.39}$$

Therefore, the analytic function Z, as opposed to V (3.29), is not harmonic.

For points on the equator of the earth, the centrifugal potential has a value of 1.1 x 10^5 m^2s^{-2}, and the centrifugal acceleration is 0.03 ms^{-2} ($\approx 0.3\%$ of gravitation). At the poles, we have $Z = 0$ and $z = 0$.

3.1.5 Gravity Acceleration, Gravity Potential

The gravity acceleration, or *gravity* $\mathbf{g}$ (Latin: gravitas), is the resultant of gravitation $\mathbf{b}$ and centrifugal acceleration $\mathbf{z}$ (Fig. 3.5):

$$\mathbf{g} = \mathbf{b} + \mathbf{z} \;. \tag{3.40}$$

By multiplying with the mass m of the attracted point, we obtain the force of gravity

$$\mathbf{F} = m\mathbf{g}.$$ (3.41)

The direction of $\mathbf{g}$ is referred to as the direction of the *plumb line* (vertical); the magnitude g is called gravity intensity (often just gravity). With (3.10) and (3.38), the *gravity potential* of the earth becomes

$$W = W(\mathbf{r}) = V + Z = G \iiint_{\text{earth}} \frac{\rho}{l} dv + \frac{\omega^2}{2} p^2.$$ (3.42)

It is related to the *gravity acceleration* by

$$\mathbf{g} = \text{grad} W.$$ (3.43)

In the X,Y,Z-system, we have

$$\mathbf{g}^{\mathsf{T}} = (\text{grad} W)^{\mathsf{T}} = (W_X, W_Y, W_Z).$$ (3.44)

Taking (2.19) into account, we obtain the components of gravity expressed by the plumb line parameters Φ, Λ:

$$\mathbf{g} = -g\mathbf{n} = -g \begin{pmatrix} \cos\Phi\cos\Lambda \\ \cos\Phi\sin\Lambda \\ \sin\Phi \end{pmatrix}.$$ (3.45)

The property

$$\text{curl} \mathbf{g} = \text{curl} \text{grad} W = 0$$ (3.46)

follows from the corresponding properties of gravitation and centrifugal acceleration and can also be expressed by the conditions

$$W_{XY} = W_{YX}, \ W_{XZ} = W_{ZX}, \ W_{YZ} = W_{ZY}.$$ (3.47)

W and its first derivatives are single-valued, finite, and continuous in the whole space as a consequence of the characteristics of V and Z. Exceptions are the uninteresting cases $r \rightarrow \infty$ (then also $Z \rightarrow \infty$) and $g = 0$ (direction of the plumb line is not unique). Due to the behavior of V, the second derivatives of W are discontinuous inside the earth at abrupt density changes. For geodesy, the most

important surface of discontinuity is the physical surface of the earth, with a density jump from 1.3 $kg\,m^{-3}$ (density of air) to 2700 $kg\,m^{-3}$ (mean density of the upper crust).

From (3.31) and (3.39), we obtain the generalized Poisson differential equation

$$\Delta W = -4\pi G\rho + 2\omega^2. \tag{3.48}$$

In outer space ($\rho = 0$), it becomes the generalized Laplace differential equation

$$\Delta W = 2\omega^2. \tag{3.49}$$

With the conditions (3.47) and (3.48) resp. (3.49), the gravity potential W possesses only five (out of nine) mutually independent second derivatives. They are closely related to the curvature of the level surfaces and the plumb lines, cf. [3.2.2].

Because of the flattening at the earth's poles and the centrifugal acceleration, g depends on the latitude. The value of gravitation for a spherical model, 9.82 ms^{-2}, cf. [3.1.2], decreases at the equator and increases at the poles of an ellipsoidal model. The centrifugal acceleration further diminishes the equatorial value, while gravitation at the poles is not influenced by centrifugal acceleration, cf. [3.1.4]. As a result, gravity varies between 9.78 ms^{-2} (equator) and 9.83 ms^{-2} (poles), see also [4.3].

3.2   Geometry of the Gravity Field

A geometrical representation of the gravity field is given by the level surfaces and the plumb lines [3.2.1]. Local field properties are described by the curvatures of level surfaces and plumb lines [3.2.2], and a system of "natural" coordinates can be based on these properties [3.2.3].

3.2.1   Level Surfaces and Plumb Lines

The surfaces of constant gravity potential

$$W = W(\mathbf{r}) = \text{const}. \tag{3.50}$$

are designated as *equipotential* or *level surfaces* (also geopotential surfaces) of gravity. As a result of an infinitesimal displacement $d\mathbf{s}$, and in view of (3.43), the potential difference of differentially separated level surfaces (Fig. 3.6) is given by
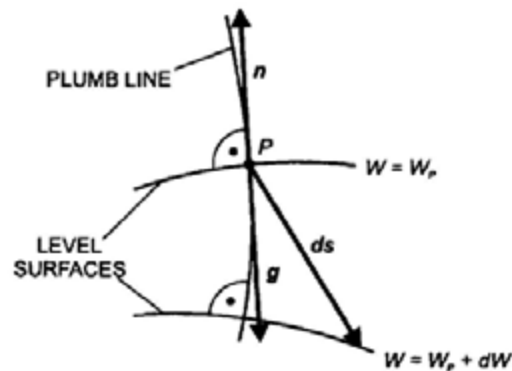
Fig. 3.6. Neighboring level surfaces

$$dW = \mathbf{g} \cdot d\mathbf{s} = g\, ds \cos(\mathbf{g}, d\mathbf{s}). \qquad (3.51)$$

This means that the derivative of the gravity potential in a certain direction is equal to the component of the gravity along this direction. Since only the projection of ds along the plumb line enters into (3.51), dW is independent of the path. Hence, no work is necessary for a displacement along a level surface $W = \text{const.}$; the level surfaces are *equilibrium* surfaces.

If ds is taken along the level surface $W = W_P$, then it follows from $dW = 0$ that $\cos(\mathbf{g}, d\mathbf{s}) = \cos 90° = 0$: gravity is normal to $W = W_P$. The level surfaces are intersected at right angles by the *plumb lines*. The tangent to the plumb line is called the *direction* of the plumb line and has been defined already in [3.1.5]. If ds is directed along the outer surface normal $\mathbf{n}$, then, because $\cos(\mathbf{g}, \mathbf{n}) = \cos 180° = -1$, the following important relationship exists:

$$dW = -g\, dn. \qquad (3.52)$$

It provides the link between the potential difference (a physical quantity) and the difference in height (a geometric quantity) of neighboring level surfaces. According to this relation, a combination of gravity measurements and (quasi) differential height determinations, as provided by geometric leveling, delivers gravity potential differences, cf. [5.5.3].

If g varies on a level surface, then, according to (3.52), the distance dn to a neighboring level surface also changes. Therefore, the level surfaces are not parallel, and the plumb lines are space curves. As a consequence of the gravity increase of $0.05\ \text{ms}^{-2}$ from the equator to the poles, the level surfaces of the earth converge toward the poles by $0.05\,\text{ms}^{-2}/9.8\ \text{ms}^{-2}$, or 5×10^{-3}, in a relative sense. Consequently, two level surfaces that are 100.0 m apart at the equator are separated by only 99.5 m at the poles (Fig. 3.7).
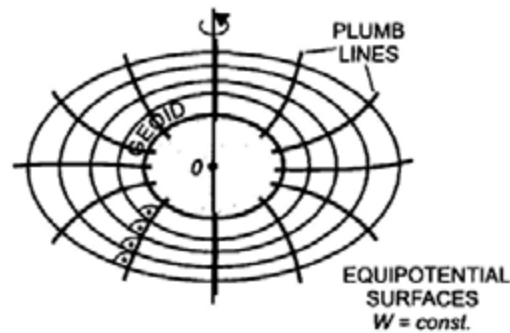
Fig. 3.7. Equipotential surfaces and plumb lines close to the earth

The level surfaces inside the earth and in the exterior space are closed spheroidal surfaces. The geoid is the level surface that approximates mean sea level. Because of its importance as a reference surface for heights, it will be treated separately in [3.4]. As an outer limit in the realm of the definition of gravity, one may consider the level surface for which the gravitation and centrifugal acceleration in the equatorial plane cancel each other. The equatorial radius of this surface would be 42 200 km.

The concept of the level surface was introduced by *MacLaurin* (1742), whereas *Clairaut* (1743) thoroughly discussed level surfaces and plumb lines as a whole. BRUNS (1878) included the determination of the exterior level surfaces in their entirety in the fundamental problem of geodesy.

3.2.2 Local Gravity Field Representation, Curvatures

From the properties of the potential function $W = W(\mathbf{r})$, it follows that the level surfaces which lie entirely in the exterior space are analytical surfaces; that is, they have no salient or singular points, cf. [3.1.5], and can be expanded in Taylor series. Level surfaces extending partially or completely inside the earth exhibit discontinuities in the second derivatives where density jumps occur. These surfaces can thus be constructed from pieces of different analytical surfaces only.

Using the local astronomic x,y,z-system introduced in [2.6.2], we develop the potential W in the vicinity of the origin P into a series. This local representation reads as

$$W = W_P + W_x + W_y + W_z + \frac{1}{2}\left(W_{xx}x^2 + W_{yy}y^2 + W_{zz}z^2\right)$$
$$+ W_{xy}xy + W_{xz}xz + W_{yz}yz + \cdots \tag{3.53}$$

Here, $W_x = \partial W / \partial x$, $W_{xx} = \partial^2 W / \partial x^2$, $W_{xy} = \partial^2 W / \partial x \partial y$, etc. represent the first and second order partial derivatives at P in the local system. If the calculation point is located on the level surface through P, we have

$$W = W_P, \; W_x = W_y = 0, \; W_z = -g.$$

By solving for z, we get the equation of the *level surface* in the neighborhood of P:

$$z = \frac{1}{2g}\left(W_{xx}x^2 + 2W_{xy}xy + W_{yy}y^2\right). \tag{3.54}$$

Here, we have neglected terms of third and higher order, taking into account that z is of second order compared to x and y, due to the small curvature of the level surfaces.

The *curvature* of the *level surface* at P along an azimuth A is described by the curvature of the normal section (intersection of the vertical plane with the surface), which is called normal curvature. The well-known formula for the depression of a sphere (local approximation to the level surface) with respect to the horizontal x,y-plane gives

$$z = -\frac{s^2}{2R_A}, \tag{3.55}$$

with s = distance from P and R_A = radius of curvature in the azimuth A (Fig. 3.8). We substitute x,y with the local polar coordinates s,A

$$x = s\cos A, \; y = s\sin A$$

and introduce (3.55) into (3.54). The *normal curvature* then reads as

$$k = \frac{1}{R_A} = -\frac{1}{g}\left(W_{xx}\cos^2 A + 2W_{xy}\sin A\cos A + W_{yy}\sin^2 A\right). \tag{3.56}$$

For the x and y-directions ($A = 0°$ and $A = 180°$), we obtain the curvatures

$$k_x = \frac{1}{R_x} = -\frac{W_{xx}}{g}, \; k_y = \frac{1}{R_y} = -\frac{W_{yy}}{g}, \tag{3.57}$$

where R_x and R_y are the corresponding curvature radii. Analogously, the geodetic torsion in the direction of the meridian (expressing the change of
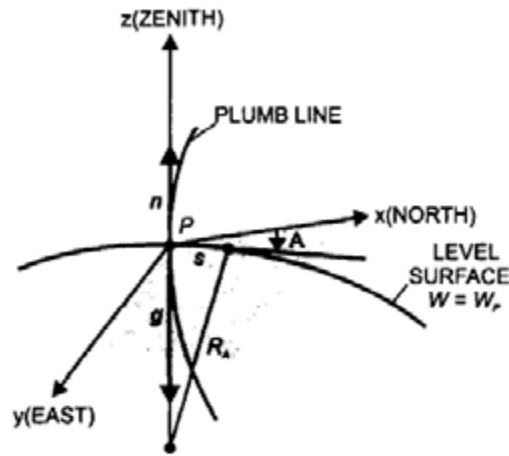
Fig. 3.8. Curvature of level surfaces and plumb lines

direction normal to the meridian) is given by

$$t_x = -\frac{W_{xy}}{g}. \tag{3.58}$$

The normal curvature assumes its extreme values in the mutually perpendicular directions of *principal curvature* A_1 and $A_2 = A_1 \pm 90°$. By considering the extrema, we find

$$\tan 2A_{1,2} = 2\frac{W_{xy}}{W_{xx} - W_{yy}}. \tag{3.59}$$

Introducing (3.59) into (3.56) yields the corresponding principal curvatures

$$\frac{1}{R_{A(1,2)}} = -\frac{1}{g}\left(W_{xx} + W_{xy}\tan A_{1,2}\right). \tag{3.60}$$

The *mean curvature* of the level surface is given by

$$J = \frac{1}{2}\left(k_x + k_y\right) = -\frac{1}{2g}\left(W_{xx} + W_{yy}\right). \tag{3.61}$$

Outside the masses of the earth, the *plumb lines* can also be described analytically. In the local astronomic system, the equation of the plumb line is given by

$$x = x(s), \quad y = y(s), \quad z = z(s), \tag{3.62}$$

where s is the arc length reckoned in the direction of gravity (Fig. 3.8). The line element along s thus differs from gravity only by the "scale factor" g:

$$g \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} W_x \\ W_y \\ W_z \end{pmatrix}, \tag{3.63}$$

with $x' = ds/dx$, etc. The curvature vector of the plumb line lies in the principal normal through P and thus in the horizontal plane. It reads as

$$\begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = \kappa \begin{pmatrix} \cos A \\ \sin A \\ 0 \end{pmatrix}, \tag{3.64}$$

where κ is the total curvature, and A is the azimuth of the principal normal. Differentiating (3.63) with respect to s, and considering that at P: $x' = y' = 0$, $z' = -1$, a substitution into (3.64) yields

$$\kappa = -\frac{W_{xz}}{g \cos A} = -\frac{W_{yz}}{g \sin A} \tag{3.65}$$

and

$$A = \arctan \frac{W_{yz}}{W_{xz}}. \tag{3.66}$$

The *curvatures* of the projections of the plumb line on the x,z-plane $(A = 0°)$ and y,z-plane $(A = 90°)$ follow from (3.65):

$$\kappa_x = -\frac{W_{xz}}{g}, \quad \kappa_y = -\frac{W_{yz}}{g}, \tag{3.67}$$

where

$$\kappa = \sqrt{\kappa_x^2 + \kappa_y^2}.$$

From (3.56) to (3.67), we recognize that the curvatures of the level surfaces and the plumb lines depend on the second derivatives of the gravity potential. Consequently, they experience discontinuities at abrupt density changes, as discussed for the potential function, cf. [3.1.5].

The *gravity gradient tensor* (*Eötvös* tensor) is comprised of the second derivatives of W as follows:

$$\left[W_{ij}\right] = \operatorname{grad} \mathbf{g} = \operatorname{grad}\left(\operatorname{grad} W\right) = \begin{pmatrix} W_{xx} & W_{xy} & W_{xz} \\ W_{yx} & W_{yy} & W_{yz} \\ W_{zx} & W_{zy} & W_{zz} \end{pmatrix}. \tag{3.68}$$

With (3.57), (3.58), (3.67), and $W_z = -g$, it can be transformed into the *Marussi* tensor

$$\left[w_{ij}\right] = -\frac{\left[W_{ij}\right]}{g} = \begin{vmatrix} k_x & t_x & \kappa_x \\ t_x & k_y & \kappa_y \\ \kappa_x & \kappa_y & \dfrac{1}{g}\dfrac{\partial g}{\partial z} \end{vmatrix}, \tag{3.69}$$

which completely describes the geometry of the gravity field (GRAFAREND 1986, MORITZ and HOFMANN-WELLENHOF 1993). As already stated in [3.1.5], (3.68) resp. (3.69) only contain five independent elements.

The Eötvös tensor (3.68) includes the *gravity gradient*

$$\operatorname{grad} g = -\begin{pmatrix} W_{xy} \\ W_{yz} \\ W_{zz} \end{pmatrix} = \begin{pmatrix} \partial g/\partial x \\ \partial g/\partial y \\ \partial g/\partial z \end{pmatrix}, \tag{3.70}$$

which describes the variation of gravity in the horizontal plane and in the vertical direction. The horizontal gradient is formed by the components $\partial g/\partial x$ and $\partial g/\partial y$, and points in the direction of maximum gravity increase in the horizontal plane. The vertical component (often called vertical gradient) $\partial g/\partial z$ describes the gravity change with height. If we combine the generalized Poisson equation (3.48) with the mean curvature (3.61), we get

$$\Delta W = W_{xx} + W_{yy} + W_{zz} = -2gJ - \frac{\partial g}{\partial z} = -4\pi G\rho + 2\omega^2$$

or

$$\frac{\partial g}{\partial z} = -2gJ + 4\pi G\rho - 2\omega^2 . \tag{3.71}$$

This relation was found by BRUNS (1878). It connects the vertical gradient with the mean curvature of the level surface and offers a possibility to determine this curvature from gravity measurements, cf. [5.4.5].

3.2.3   Natural Coordinates

We introduce a system of non-linear "natural" coordinates Φ, Λ, W defined in the gravity field. Astronomical latitude Φ and astronomical longitude Λ describe the direction of the plumb line at the point P. They have been introduced already in [2.6.1] as orientation parameters of the local gravity field system with respect to the global geocentric system. The gravity potential W locates P in the system of level surfaces $W = $ const. (Fig. 2.12). Hence, P is determined by the non-orthogonal intersection of the coordinate surfaces $\Phi = $ const., $\Lambda = $ const., and the surface $W = $ const. The coordinate lines (spatial curves) are called astronomic meridian curve ($\Lambda, W = $ const.), astronomic parallel curve ($\Phi, W = $ const.), and isozenithal line ($\Phi, \Lambda = $ const.).

The natural coordinates can be determined by measurements. Astronomic positioning provides latitude and longitude, cf. [5.3.2]. Although W cannot be measured directly, potential differences can be derived from leveling and gravity measurements and then referred to a selected level surface, e.g., the geoid, cf. [5.5.3].

The relationship between the global X, Y, Z-system and the Φ, Λ, W-system is obtained from (3.45):

$$\mathbf{g} = grad\, W = -g \begin{pmatrix} \cos\Phi\cos\Lambda \\ \cos\Phi\sin\Lambda \\ \sin\Phi \end{pmatrix} . \tag{3.72}$$

Solving for the natural coordinates yields the highly non-linear relations:

$$\left. \begin{aligned} \Phi &= \arctan\frac{-W_z}{\sqrt{W_x^2+W_y^2}} \\[2mm] \Lambda &= \arctan\frac{W_Y}{W_X} \\[2mm] W &= W(X,Y,Z) \end{aligned} \right\}. \qquad (3.73)$$

Differential relations between the *local* Cartesian coordinates x,y,z (local astronomic system) and the *global* Φ,Λ,W-system are given by

$$d\Phi = \frac{\partial \Phi}{\partial x}dx + \frac{\partial \Phi}{\partial y}dy + \frac{\partial \Phi}{\partial z}dz \text{ , etc.,}$$

where dx, dy, and dz can be derived from local measurements, cf. [2.6.2].

The partial derivatives of Φ and Λ describe the change of the plumb line direction when moving in the gravity field. This corresponds to the curvature of the level surface (when moving in the horizontal plane) and of the plumb line (when moving vertically). We have the following relations:

$$\left. \begin{aligned} &\frac{\partial \Phi}{\partial x}=k_x, \quad \frac{\partial \Phi}{\partial y}=\frac{\cos\Phi\,\partial\Lambda}{\partial x}=t_x, \quad \frac{\partial \Phi}{\partial z}=\kappa_x \\[2mm] &\frac{\cos\Phi\,\partial\Lambda}{\partial y}=k_y, \quad \frac{\cos\Phi\,\partial\Lambda}{\partial z}=\kappa_y \\[2mm] &\frac{\partial W}{\partial x}=0, \quad \frac{\partial W}{\partial y}=0, \quad \frac{\partial W}{\partial z}=-g \end{aligned} \right\}, \qquad (3.74)$$

where the curvature and torsion parameters are given by (3.57), (3.58), and (3.67). Introducing (3.74) into the differential relations leads to the transformation

$$\begin{pmatrix} d\Phi \\ \cos\Phi\,d\Lambda \\ dW \end{pmatrix} = \begin{pmatrix} k_x & t_x & \kappa_x \\ t_x & k_y & \kappa_y \\ 0 & 0 & -g \end{pmatrix} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix}, \qquad (3.75)$$

which again contains the elements of the Marussi tensor (3.69), see GRAFAREND (1972), MORITZ and HOFMANN-WELLENHOF (1993).

As the orientation of the local systems changes from point to point, the differentials dx, dy, dz are imperfect ones, with loop closures differing from zero:

$$\oint dx \neq 0, \quad \oint dy \neq 0, \quad \oint dz \neq 0. \qquad (3.76)$$

Φ, Λ, W, on the other hand, possess perfect differentials with

$$\oint d\Phi = 0, \quad \oint d\Lambda = 0, \quad \oint dW = 0. \qquad (3.77)$$

Equation (3.75) offers the possibility to transform local observable-quantities (azimuths, horizontal directions and angles, zenith angles, distances, potential differences from leveling and gravity) to the global gravity field system, where the astronomic latitude and longitude coordinates are also observable.

A theory of "intrinsic geodesy" based on the differential geometry of the gravity field has been developed by MARUSSI (1949,1985), see also HOTINE (1969). Using only observable quantities, reductions to conventional reference systems are completely avoided. On the other hand, in order to practically evaluate (3.75), a detailed knowledge of the curvature of the gravity field would be necessary. This would require a dense survey of the second derivatives of the gravity potential, as the curvature close to the earth's surface is rather irregular. Present gravity models do not sufficiently provide this information, and gravity gradiometer techniques are either time consuming or not yet operational, cf. [5.4.5]. But even with a better knowledge of the curvature of the gravity field, the transfer of coordinates would hardly be made in the system of natural coordinates, due to the complex structure of the gravity field.

3.3   Spherical Harmonic Expansion of the Gravitational Potential

Because the density function $\rho = \rho(r')$ of the earth is not well known, the gravitational potential $V = V(r)$ cannot be computed by Newton's law of gravitation using (3.10). However, a convergent series expansion of V is possible in the exterior space as a special solution of Laplace's differential equation (3.29). It can be easily derived from an expansion of the reciprocal distance appearing in Newton's law [3.3.1], [3.3.2], see KELLOGG (1929), SIGL (1985), BLAKELEY (1996). This solution corresponds to a spectral decomposition of the gravitational field [3.3.3]. The coefficients of the series expansion provide the amplitudes of the respective spectral parts [3.3.4]. Any observable functional of V can be evaluated for the determination of these coefficients, thus allowing a global analytical representation of the gravitational field, cf. [6.6.1].

3.3.1   Expansion of the Reciprocal Distance

Applying the law of cosines to the triangle $OP'P$ (Fig. 3.1), we obtain

$$\frac{1}{l}=\left(r^2+r'^2-2rr'\cos\psi\right)^{-\frac{1}{2}}=\frac{1}{r}\left(1+\left(\frac{r'}{r}\right)^2-2\frac{r'}{r}\cos\psi\right)^{-\frac{1}{2}} \qquad (3.78)$$

for the reciprocal distance $1/l$ appearing in (3.10), between the attracted point P and the attracting point P'. Here, ψ is the central angle between the directions from O to P and O to P' respectively. If $1/l$ is expanded in a series converging for $r'<r$, and if the terms are arranged according to increasing powers of r'/r, then it follows that

$$\frac{1}{l}=\frac{1}{r}\sum_{l=0}^{\infty}\left(\frac{r'}{r}\right)^l P_l(\cos\psi). \qquad (3.79)$$

The $P_l(\cos\psi)$ terms represent polynomials of lth degree in $\cos\psi$. They are known as *Legendre polynomials* (zonal harmonics), and they are computed for the argument $t=\cos\psi$ by means of

$$P_l(t)=\frac{1}{2^l\times l!}\times\frac{d^l}{dt^l}\left(t^2-1\right)^l. \qquad (3.80a)$$

A rapid calculation is possible with the recurrence formula (WENZEL 1985)

$$P_l(t)=\frac{2l-1}{l}t P_{l-1}(t)-\frac{l-1}{l}P_{l-2}(t),\; l\geq 2, \qquad (3.80b)$$

with $P_0=1$, $P_1=t$.

We now introduce the unit sphere σ around the origin of the coordinates O (Fig. 3.9). The projections of OP and OP' on σ, together with the north pole projection N, form a spherical triangle. It contains the spherical coordinates ϑ,λ and ϑ',λ', and the central angle ψ appears as a spherical distance on σ, see also [2.5.1]. Spherical trigonometry provides the following relationship

$$\cos\psi=\cos\vartheta\cos\vartheta'+\sin\vartheta\sin\vartheta'\cos(\lambda'-\lambda).$$

The corresponding decomposition of $P_l(\cos\psi)$ leads to
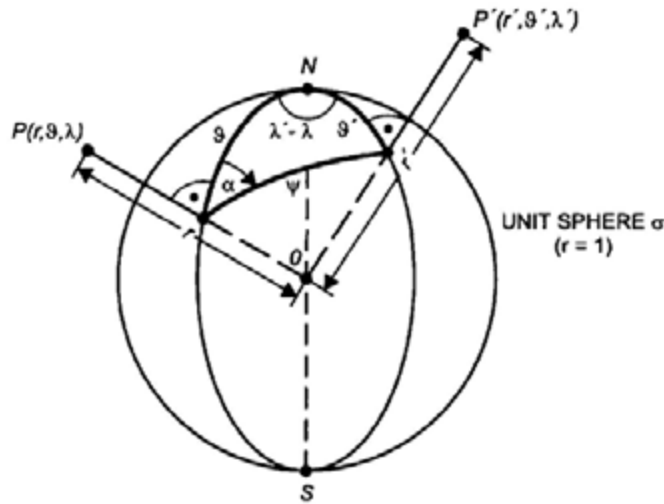
Fig. 3.9. Spherical polar triangle on the unit sphere, spherical coordinates

$$P_l\left(\cos\psi\right)= P_l\left(\cos\vartheta\right)P_l\left(\cos\vartheta'\right)$$

$$+2\sum_{m=1}^{l}\frac{(l-m)!}{(l+m)!}(P_{lm}\left(\cos\vartheta\right)\cos m\lambda\,P_{lm}\left(\cos\vartheta'\right)\cos m\lambda' \quad (3.81)$$

$$+P_{lm}\left(\cos\vartheta\right)\sin m\lambda\,P_{lm}\left(\cos\vartheta'\right)\sin m\lambda').$$

Again, the $P_l(t)$ are the Legendre polynomials with the argument $t=\cos\vartheta$ or $t=\cos\vartheta'$. The *associated* Legendre functions of the first kind, $P_{lm}(t)$ ($l =$ degree, $m =$ order), are obtained by differentiating $P_l(t)$ m times with respect to t:

$$P_{lm}\left(t\right)=\left(1-t^2\right)^{\frac{m}{2}}\frac{d^m}{dt^m}P_l\left(t\right). \quad (3.82)$$

Up to degree 3, the Legendre polynomials and the associated Legendre functions are given as follows:

$$P_0 =1,\ P_1 =\cos\vartheta,\ P_2 =\frac{3}{2}\cos^2\vartheta-\frac{1}{2},\ P_3 =\frac{5}{2}\cos^3\vartheta-\frac{3}{2}\cos\vartheta \quad (3.83a)$$

and

$$P_{1,1} = \sin\vartheta, \quad P_{2,1} = 3\sin\vartheta\cos\vartheta, \quad P_{2,2} = 3\sin^2\vartheta$$

$$P_{3,1} = \sin\vartheta\left(\frac{15}{2}\cos^2\vartheta - \frac{3}{2}\right), \quad P_{3,2} = 15\sin^2\vartheta\cos\vartheta, \quad P_{3,3} = 15\sin^3\vartheta \quad .(3.83b)$$

A series development for the calculation of $P_{lm}(t)$ is given in HEISKANEN and MORITZ (1967, p.24). By substituting (3.81) into (3.79), the expansion of $1/l$ into spherical harmonics is completed.

The functions

$$Y_{lm}^c(\vartheta,\lambda) = P_{lm}(\cos\vartheta)\cos m\lambda$$
$$Y_{lm}^s(\vartheta,\lambda) = P_{lm}(\cos\vartheta)\sin m\lambda \quad , \tag{3.84}$$

which appear in (3.81), are called *Laplace's surface spherical harmonics*. They characterize the behavior of the developed function (here $1/l$) on the unit sphere, cf. [3.3.3]. The *orthogonality* relations are valid for these functions, i.e., the integral over the unit sphere of the product of any two different functions is zero:

$$\iint_\sigma Y_{lm}^i Y_{nq}^k d\sigma = 0 \tag{3.85a}$$

for $n \neq l$, $q \neq m$, or $k \neq i$. For the product of two equal functions Y_{lm}^c or Y_{lm}^s, we have

$$\iint_\sigma Y_{lm}^2 d\sigma = \begin{cases} \dfrac{4\pi}{2l+1} & for \ m = 0 \\[2ex] \dfrac{2\pi}{2l+1}\dfrac{(l+m)!}{(l-m)!} & for \ m \neq 0 \end{cases}, \tag{3.85b}$$

see HEISKANEN and MORITZ (1967, p.29).

3.3.2 Expansion of the Gravitational Potential

We substitute the spherical harmonic expansion of $1/l$, (3.79) and (3.81), into the volume integral (3.10):

$$V = \frac{G}{r} \sum_{l=0}^{\bar{\infty}} \sum_{m=0}^{l} k \frac{(l-m)!}{(l+m)!} \times$$

$$\frac{1}{r'} \left(\begin{array}{l} P_{lm}(\cos\vartheta)\cos m\lambda \iiint\limits_{earth} r'^l P_{lm}(\cos\vartheta')\cos m\lambda' \, dm \\ +P_{lm}(\cos\vartheta)\sin m\lambda \iiint\limits_{earth} r'^l P_{lm}(\cos\vartheta')\sin m\lambda' \, dm \end{array} \right). \qquad (3.87)$$

$$k = \begin{cases} 1 & \text{for } m = 0 \\ 2 & \text{for } m \neq 0 \end{cases}$$

In abbreviated form this development can be expressed as

$$V = \sum_{l=0}^{\bar{\infty}} V_l = \sum_{l=0}^{\bar{\infty}} \frac{Y_l(\vartheta,\lambda)}{r^{l+1}}, \qquad (3.88)$$

where the V_l are called *solid spherical harmonics*, and the $Y_l(\vartheta,\lambda)$ are linear combinations of the surface spherical harmonics (3.84).

For $l = 0$, the integration yields the potential of the earth's mass M concentrated at the center of mass (3.16). We extract this term, introduce the semimajor axis a of the earth ellipsoid as a constant, and denote the mass integrals by C_{lm}, S_{lm} (spherical harmonic coefficients). The gravitational potential expanded in spherical harmonics is then written as

$$V = \frac{GM}{r} \left(1 + \sum_{l=1}^{\bar{\infty}} \sum_{m=0}^{l} \left(\frac{a}{r} \right)^l (C_{lm} \cos m\lambda + S_{lm} \sin m\lambda) P_{lm}(\cos\vartheta) \right). \quad (3.89)$$

The *harmonic coefficients* (also Stokes' constants) are given by

$$\left. \begin{array}{l} C_{l0} = C_l = \frac{1}{M} \iiint\limits_{earth} \left(\frac{r'}{a} \right)^l P_l(\cos\vartheta') dm \qquad \text{and} \\ \\ \begin{Bmatrix} C_{lm} \\ S_{lm} \end{Bmatrix} = \frac{2}{M} \times \frac{(l-m)!}{(l+m)!} \iiint\limits_{earth} \left(\frac{r'}{a} \right)^l P_{lm}(\cos\vartheta') \begin{Bmatrix} \cos m\lambda' \\ \sin m\lambda' \end{Bmatrix} dm \quad \text{for } m \neq 0 \end{array} \right\} . (3.90)$$

The following coefficients are sometimes used, particularly in satellite geodesy:

$$J_l = -C_l, \quad J_{lm} = -C_{lm}, \quad K_{lm} = -S_{lm}. \qquad (3.91)$$

Calculations in the gravitational field become more convenient with the (fully) *normalized* spherical harmonic functions $\bar{P}_{lm}(\cos\vartheta)$. They are computed from the conventional harmonics, (3.80) and (3.82), according to

$$\bar{P}_{lm}(t) = \sqrt{k(2l+1)\frac{(l-m)!}{(l+m)!}} P_{lm}(t), \quad k = \begin{cases} 1 & for\ m=0 \\ 2 & for\ m\neq 0 \end{cases}, \quad (3.92)$$

with $t = \cos\vartheta$, etc. Recursive formulas are also available for the calculation of the normalized harmonics and its derivatives (PAUL 1978, TSCHERNING et al. 1983, WENZEL 1985):

$$\bar{P}_{lm}(t) = \left[\frac{(2l+1)(2l-1)}{(l+m)(l-m)}\right]^{1/2} t\bar{P}_{l-1,m}(t) -$$

$$\left[\frac{(2l+1)(l+m-1)(l-m-1)}{(2l-3)(l+m)(l-m)}\right]^{1/2} \bar{P}_{l-2,m}(t) \quad (3.93a)$$

for $l > m+1$

with

$$\bar{P}_0 = 1, \quad \bar{P}_1 = \sqrt{3}\cos\vartheta, \quad \bar{P}_2 = \frac{1}{2}\sqrt{5}(3\cos^2\vartheta - 1),$$

$$(3.93b)$$

$$\bar{P}_{1,1} = \sqrt{3}\sin\vartheta, \quad \bar{P}_{2,1} = \sqrt{15}\sin\vartheta\cos\vartheta, \quad \bar{P}_{2,2} = \frac{1}{2}\sqrt{15}\sin^2\vartheta$$

and the control

$$\sum_{m=0}^{l} \bar{P}_{lm}(t)^2 = 2l+1. \quad (3.93c)$$

In addition to the orthogonality relation (3.85) for the surface harmonics, we now also have

$$\frac{1}{4\pi}\iint_{\sigma}\left(\bar{P}_{lm}\begin{Bmatrix}\cos m\lambda \\ \sin m\lambda\end{Bmatrix}\right)^2 d\sigma = 1 \quad (3.94)$$

over the unit sphere σ. According to (3.92), for an expansion of the gravitational potential analogous to (3.89), the harmonic coefficients are now given by

$$\begin{Bmatrix} \overline{C}_{lm} \\ \overline{S}_{lm} \end{Bmatrix} = \sqrt{\frac{(l+m)!}{k(2l+1)(l-m)!}} \begin{Bmatrix} C_{lm} \\ S_{lm} \end{Bmatrix}, \quad k = \begin{cases} 1 & for\ m=0 \\ 2 & for\ m \neq 0 \end{cases} . \tag{3.95}$$

Equation (3.89), or the corresponding equations employing fully normalized harmonics ((3.92) and (3.95)), represent a spherical solution of Laplace's differential equation (3.29). It can also be derived straightforwardly by the method of variable separation, after substituting the Cartesian coordinates with spherical coordinates (HEISKANEN and MORITZ 1967, pp.18 ff.).

The expansion converges outside a sphere of radius $r = a$, which just encloses the earth (Brillouin sphere). After the theorem of Runge-Krarup, an expansion of V into converging spherical harmonics can also be used in the interior of the earth, down to a sphere completely inside the earth and close to its surface (Bjerhammar sphere), KRARUP (1969), MORITZ (1980, p.69). Such an expansion represents an arbitrarily good approximation of the outer gravitational field. Naturally, this analytical extension does not satisfy the Poisson equation (3.31) of the actual gravitational field in the earth's interior.

With present accuracies of the determination of the earth's gravity field, the gravitation of the *atmosphere* cannot be neglected. As the density of the atmosphere primarily depends on height, corresponding models can be used to calculate the potential and the gravitation of the atmosphere as a function of height. These calculations are based on the relations derived for the potential inside an earth constructed of homogeneous spherical shells, cf. [3.1.2]. With an atmospheric mass of 5.32×10^{18} kg, we get a potential value of 55.6 m^{2}s^{-2} for $h = 0$, and 54.8 m^{2}s^{-2} for $h = 100$ km. This effect is taken into account by corresponding reductions, cf. [4.3].

The extension of the spherical harmonic expansion for V to the *gravity potential W* is performed easily by adding the centrifugal potential Z (3.38). If we express the distance p to the rotational axis by spherical coordinates (2.14), the centrifugal potential reads as

$$Z = \frac{\omega^2}{2} r^2 \sin^2 \vartheta \tag{3.96a}$$

or, after introducing the Legendre polynomial P_2 according to (3.83a), as

$$Z = \frac{\omega^2}{3} r^2 \left(1 - P_2(\cos\vartheta)\right). \tag{3.96b}$$

By adding (3.96) to (3.89) we get the expansion for the gravity potential.

3.3.3 Geometrical Interpretation of the Surface Spherical Harmonics

We now discuss the properties of the surface spherical harmonics (3.84), which describe the behavior of the gravitational potential on the unit sphere. The zero points of these functions divide the surface into regions with alternating signs, bounded by meridians and parallels.

For the order $m = 0$, we obtain the Legendre polynomials $P_l(\cos\vartheta)$. Because of their independence of the geographical longitude λ, they divide the surface into zones of positive and negative signs: *zonal harmonics*. These harmonics possess l real zeros in the interval $0 \leq \vartheta \leq \pi$. For even l, the sphere is divided symmetrically with respect to the equator $\vartheta = 90°$, and the case for odd l results in an asymmetric division. The $P_{lm}(\cos\vartheta)$ for $m \neq 0$ have $(l-m)$ zeros in the interval $0 < \vartheta < \pi$. Because of the multiplication by $\cos m\lambda$ or $\sin m\lambda$, the surface harmonics are longitude dependent, furnishing $2m$ zeros in the interval $0 \leq \lambda \leq \pi$: *tesseral harmonics*. Finally, for $m = l$, the dependence on ϑ disappears, and the sphere is divided into sectors of alternating signs: *sectorial harmonics* (Fig. 3.10).
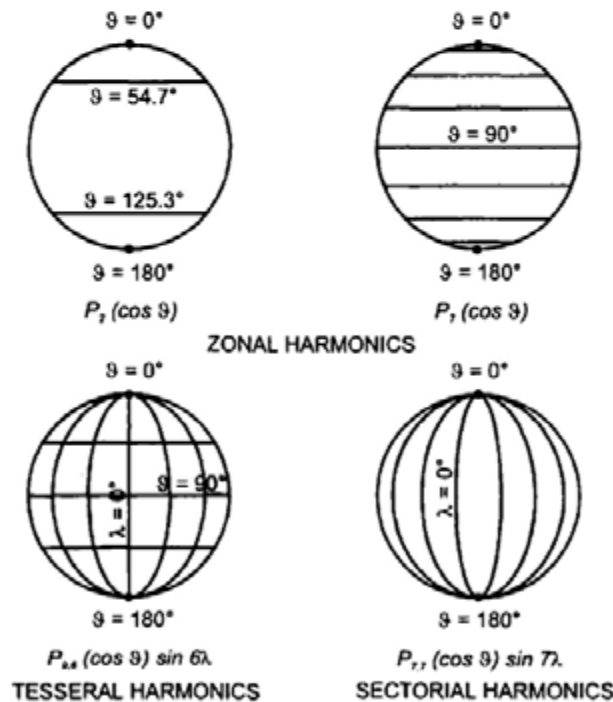
Fig. 3.10. Spherical harmonics on the unit sphere, with alternating positive (grey) and negative (white) sign

The amplitudes of the individual terms given by the surface harmonics are determined by the harmonic coefficients. For example, the series has only zonal harmonics for an earth rotationally symmetric with respect to the Z-axis; the coefficients with $m \neq 0$ must all vanish. For a mass distribution symmetric with respect to the equator, the zonal harmonic coefficients with odd l must be absent.

Summarizing, we state that the spherical harmonic expansion of the gravitational potential represents a *spectral decomposition* of the gravitational field. The field is separated into structures of wave-length $360°/l$, corresponding to a spatial resolution of $180°/l$. With increasing height, the field is smoothed by the factor $(a/r)^l$.

3.3.4   Physical Interpretation of the Harmonic Coefficients

The spherical harmonic expansion has transformed the single volume integral over the earth's masses (3.10) into an infinite series. The harmonic coefficients now carry mass integrals for the individual contribution of the corresponding wave-length to the total potential. The lower degree harmonics have a simple physical interpretation.

As already stated above, the zero degree term $(l=0)$ represents the potential of a homogeneous or radially layered spherical earth, see (3.16):

$$V_0 = \frac{GM}{r}.\qquad\qquad(3.97)$$

The terms of degree one and two $(l=1,2)$ can be calculated from (3.90) by introducing the harmonic functions P_{lm} (3.83) and subsequently transforming the spherical coordinates to Cartesian coordinates using (2.14). For $l=1$, this yields

$$C_1 = \frac{1}{aM}\iiint_{\text{earth}} Z'\,dm,\quad C_{1,1} = \frac{1}{aM}\iiint_{\text{earth}} X'\,dm,\text{ and } S_{1,1} = \frac{1}{aM}\iiint_{\text{earth}} Y'\,dm.\ (3.98)$$

As known from mechanics, the integrals divided by the mass M are the coordinates of the center of mass of the earth. Since we have placed the origin of the coordinate system at the center of mass, we have

$$C_1 = C_{1,1} = S_{1,1} = 0.\qquad\qquad(3.99)$$

For $l=2$, we obtain

$$C_2 = \frac{1}{a^2 M} \iiint\limits_{earth} \left(Z'^2 - \frac{X'^2 + Y'^2}{2} \right) dm, \quad C_{2,1} = \frac{1}{a^2 M} \iiint\limits_{earth} X'Z' dm,$$

$$S_{2,1} = \frac{1}{a^2 M} \iiint\limits_{earth} Y'Z' dm, \quad C_{2,2} = \frac{1}{4a^2 M} \iiint\limits_{earth} (X'^2 - Y'^2) dm, \quad (3.100)$$

$$S_{2,2} = \frac{1}{2a^2 M} \iiint\limits_{earth} X'Y' dm.$$

These expressions are functions of the moments of inertia

$$A = \iiint (Y'^2 + Z'^2) dm, \quad B = \iiint (X'^2 + Z'^2) dm,$$

$$C = \iiint (X'^2 + Y'^2) dm \quad (3.101)$$

and of the products of inertia

$$D = \iiint Y'Z' dm, \quad E = \iiint X'Z' dm, \quad F = \iiint X'Y' dm \quad (3.102)$$

with respect to the axes of the global X,Y,Z-system. If we neglect polar motion, the Z-axis coincides with one principal axis of inertia (maximum moment of inertia). Consequently, we have

$$D = E = 0.$$

F, on the other hand, would only become zero if the X-axis coincided with one of the equatorial principal-axes of inertia. Due to the conventional definition of the X-axis (Greenwich meridian), F does not vanish.

Using the above expressions for A, B, C, and F, the harmonic coefficients of second degree may also be formulated as follows:

$$C_2 = \frac{1}{a^2 M} \left(\frac{A+B}{2} - C \right), \quad C_{2,1} = S_{2,1} = 0,$$

$$C_{2,2} = \frac{B-A}{4a^2 M}, \quad S_{2,2} = \frac{F}{2a^2 M} \quad (3.102)$$

$J_2 = -C_2$ is also known as *dynamical form factor*. It characterizes the polar flattening of the earth's body by the difference between the mean equatorial moment of inertia (with $A \approx B$) and the polar moment of inertia. As this is the largest deviation from a spherical earth model, the numerical value for C_2 is

three orders of magnitude larger than the values of the successive coefficients. Slight contributions to the ellipsoidal form of the earth also come from the even zonal harmonics of higher degree, mainly $l = 4$ and $l = 6$. The coefficients $C_{2,2}$ and $S_{2,2}$ describe the asymmetry of the equatorial mass distribution in relation to the rotational axis (ellipticity of the equator) and the torsion of the corresponding principal axes of inertia with respect to the conventional X and Y-directions, cf. also [4.2.1]. If the odd zonal harmonic-coefficients differ from zero, the corresponding terms in the expansion of V represent an asymmetric mass distribution with respect to the equatorial plane, cf. [3.3.3]. The main contribution comes from C_3, and may be geometrically interpreted as a difference in the flattening for the northern and the southern hemisphere. Numerical values for the coefficients are given in [6.6.2].

3.4   The Geoid

The geoid is of fundamental importance for geodesy, oceanography, and physics of the solid earth. Due to the demands for, and achievable accuracy of, a modern geoid, the classical definition of the geoid must be reconsidered today [3.4.1]. In geodesy and oceanography, the geoid serves as a height reference surface for describing continental and sea surface topography [3.4.2], [3.4.3]. Geophysics exploits the geoid as a gravity field representation revealing the distribution of deeper located masses, cf. [8.2.4].

3.4.1   Definition

The geoid has been introduced by C. F. Gauss as a refined model of the figure of the earth, cf. [1.2]. It was defined by Gauss as the equipotential surface of the earth's gravity field coinciding with the mean sea level of the oceans.

This physical *definition* considers the waters of the oceans as freely-moving homogeneous matter, subject only to the force of gravity and free from variations with time. Upon attaining a state of equilibrium, the surface of such idealized oceans would assume a level surface of the gravity field. The ocean surface may be regarded as being extended under the continents, e.g., by a system of conducting tubes.

With the gravity potential value W_0, the *equation* of the geoid reads as

$$W = W(\mathbf{r}) = W_0. \tag{3.103}$$

It follows from the properties of the gravity potential W, that the geoid is a closed and continuous surface, cf. [3.1.3]. As it extends partially inside the solid

earth (under the continents), its curvature will display discontinuities at abrupt density changes. Thus, although not being an analytical surface in a global sense, it may be sufficiently well approximated by a spherical harmonic development, cf. [3.3.2].

The problem of downward continuation of the gravity field into the earth's masses presupposes knowledge about the density distribution of the atmospheric and topographic masses. Geoid calculations consequently depend on the corresponding assumptions (STRYKOWSKI 1998), cf. [6.5.3].

As well known from oceanography, mean sea level is not an equilibrium surface in the earth's gravity field, due to ocean currents and other quasi-stationary effects. In addition, sea surface variations with time can only be partially reduced by averaging over time or modeling. Hence, mean sea level still varies over longer time spans, cf. [3.4.2].

Consequently, a *refined* geoid *definition* is needed at the "cm" accuracy level (RIZOS 1982, RAPP 1983a). By applying a minimum condition on the deviations between mean sea level and the geoid, the geoid could be defined as the equipotential surface which best fits mean sea level at a certain epoch (MATHER 1978, RAPP 1995a).

Another choice would be to define the geoid as the level surface which optimally fits mean sea level at a selected set of tide gauges used for defining the vertical datum of national or continental height systems, cf. [3.4.3]. Such a definition would lead to only small corrections for the existing height systems but not result in a best fit over the oceans.

The geoid definition also has to include the treatment of the *permanent tidal effects*, cf. [3.5.2]. There are three different definitions possible. The *mean geoid* includes the direct effect of attraction and the indirect effect of deformation. It would coincide with an "undisturbed" sea level; hence it is of interest for oceanography. For the *non-tidal geoid*, the total tidal-effect would be eliminated. This would agree with the theoretical demand of geodesy to have no masses outside the boundary surface "geoid", cf. [6.5.3]. As the response of the earth to the permanent tidal part is not known, the *zero-geoid* is preferred in geodesy. Here, the attraction part is eliminated but the permanent deformation retained. This definition takes into account the fact that positioning also refers to a deformed earth (IAG resolution, Gen. Ass. Hamburg 1983, EKMAN 1989). With respect to a best-fitting reference ellipsoid, cf. [4.3], the geoid deviates by ±30 m; maximum deviations reach about ±100 m, cf. [6.6.3].

Finally, a refined geoid definition should also consider *variations with time*, which result from the displacements of terrestrial masses, cf. [3.5.3]. Such a definition then would refer to a certain epoch.

3.4.2   Mean Sea Level

The ocean surface does not coincide with a level surface (e.g. the geoid) of the earth's gravity field; the deviations are called *sea surface topography* (SST, also dynamic ocean topography). *Instantaneous* SST is affected by temporal variations of long-term, annual, seasonal, and short-term character, occurring at different scales. Averaging the ocean surface over time (at least over one year) or modeling ocean tides provides *mean sea level* (MSL) for the corresponding time interval. Even after reducing all time-dependent parts, a quasi-stationary SST would remain. It is caused by nearly constant oceanographic and meteorological effects, which generate ocean currents and ocean surface slopes. The r.m.s. variation is ±0.6 to ±0.7 m , and the maximum deviation from the geoid is about ±1 m or more (LISITZIN 1974).

Short term variations of the sea surface (waves) are averaged out in the mean value over time (e.g., at tide gauge observations over one hour) or by the smoothing effect of the "footprints" in satellite altimetry, cf. [5.2.7]. *Tidal effects* can deviate considerably between the open ocean and shelf areas, adjacent seas, and coastal zones. This is due to unequal water depths and to the fact that the continents impede the movement of water. On the open sea, the tidal amplitude is less than one meter (r.m.s. variation ±0.3 m), while it can amount to several meters in coastal areas (Bay of Fundy, Nova Scotia: more than 15 m).

Oceanic tidal models are based on Laplace's tidal equations. These models take bathymetry, tide gauge data, and in more recent solutions, satellite altimetry and satellite orbit analyses into account. The models solve for about 10 to 12 partial tides (annual, semiannual, monthly, fortnightly, daily, and half-daily). They are provided either in grid form (1°×1° , 0.5°×0.5°) or as a spherical harmonic expansion (SCHWIDERSKI 1980, 1983, ANDERSEN et al. 1995). Ocean tide models have also been derived by global gravity modeling (SCHWINTZER et al. 1997, LEMOINE et al. 1998). The accuracy of the oceanic tidal models amounts to a few cm on the open oceans but is less at shelf areas and close to the coast.

Fluctuations of annual, semiannual, and seasonal character are of *meteorological* origin (atmospheric pressure, winds), of *oceanographic* nature (ocean currents, differences in water density as a function of temperature, salinity and pressure), and are also due to a variable *water budget* (changing water influx as a result of meltwater, monsoon rains, etc.). The amplitude of these variations is on the order of 0.1 to 1 m, and scales of a few 100 km are found, e.g., at meandering ocean streams and eddies (NEREM 1995b). In addition, a *secular rise* of about 1 to 2 mm/a has been observed over the last 100 years. This trend is expected to increase, reaching 0.5 to 1 m over the next century, due to climate changes producing a thermal expansion of the water masses, a melting of the polar ice caps and the glaciers, and isostatic movements (LAMBECK 1988, HANNAH 1989).

Mean sea level can be derived from tide gauge records, satellite altimetry, and oceanographic methods.

Tide gauges (mareographs) continuously record the height of the water level with respect to a height reference surface close to the geoid, cf. [3.4.3]. Averaging the results over long time intervals (month, year) eliminates most variations with time. In order to fully remove the tidal period of a complete lunar cycle (nutation), the record should extend over 18.6 years, cf. [2.4.2]. The precision of the mean monthly and annual values is better than ± 1 cm. These results may be systematically disturbed if the tide gauge location is not directly linked to the open ocean and data is thereby affected by local sea level anomalies (swell in shallow waters, estuary effects at river mouths). In addition, local or regional vertical crustal movements (sedimentary subsidence, postglacial uplift, etc.) may act at the tide gauge location and systematically influence (bias) the sea level registration. These movements may reach a few mm/a (EKMAN 1993, MITCHUM 1994).

Tide gauge data are available for over 1750 stations worldwide, but only a few stations cover a time span of a few centuries (at Amsterdam registrations go back to 1700), WOODWORTH (1997). In the open oceans, pressure tide gauges contribute in monitoring sea surface variability but lack connection to continental height systems (TOLKATCHEV 1996). Tide gauges along the continental coasts generally have been connected to the geodetic height control system, thus permitting the derivation of MSL from the geoid or a zero height reference close to it. Sea level slopes up to several 0.1 m/1000 km and more have been detected by this method, cf. [3.4.3].

Satellite altimetry directly provides sea surface topography with respect to an ellipsoidal reference surface, cf. [5.2.7]. With the exception of the polar regions, satellite altimeters cover the oceans with repeated tracks (e.g., with a 10 days repetition rate at TOPEX/POSEIDON) and permit derivation of mean sea surface heights to an accuracy of a few cm. The solutions are derived over a certain time period (e.g., one year) and are provided in grids of a few minutes of arc (ANZENHOFER et al. 1996, RAPP and YI 1997). Sea surface topography is obtained by referring these results to a geoid model. If the altimetric solutions for different epochs are compared, sea surface variations with time can be determined, MINSTER et al. 1995, cf. [8.3.2].

Oceanographic methods derive sea surface topography from measurements at sea (LAMBECK 1988, RUMMEL and ILK 1995). *Steric leveling* assumes that equipotential and isobaric surfaces coincide at a certain depth (e.g., 2000 m): level of no motion. Using water density values derived from salinity, temperature, and pressure data, the equation of hydrostatic equilibrium yields the gravity potential difference (or the dynamic height) between the ocean surface and the reference level of no motion. This method is applicable in the deep oceans (STURGES 1974). *Geostrophic leveling* (dynamic leveling) is based on the hydrodynamic equations and uses observed ocean current velocities. It can also be applied in shelf areas.
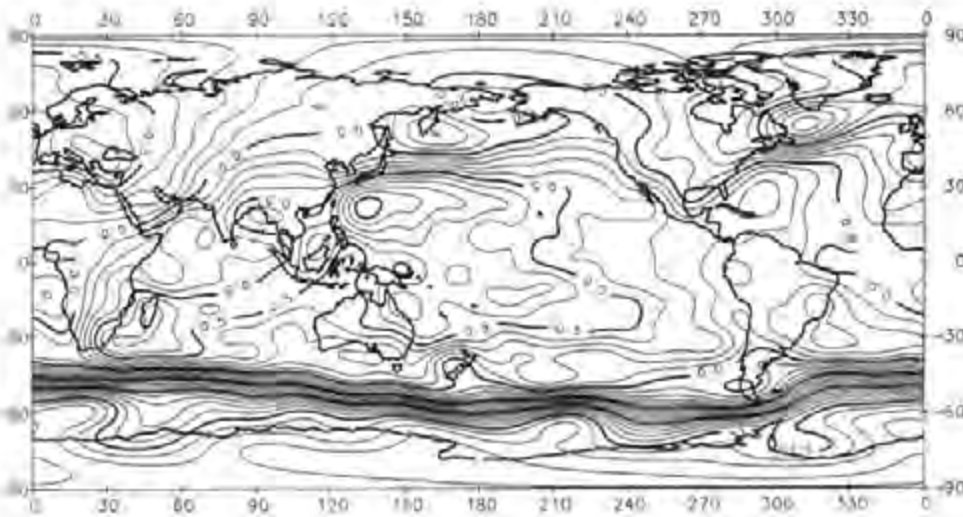
Fig. 3.11. Sea surface topography, spherical harmonic expansion up to degree 20, contour line interval 0.1 m, after LEMOINE et al. (1998)

Oceanic sea surface topography models have been developed that correspond to a spherical harmonic expansion up to degree 36 (or a minimum wave-length of 10°) with an accuracy of a few cm to 0.1 m (LEVITUS 1982). Satellite altimetry solutions, partly in combination with gravity field modeling, are available up to degree 20, for time spans over 5 to 10 years, and with an accuracy of better than 0.1 m (SCHWINTZER et al. 1997, LEMOINE et al. 1998), Fig 3.11. Quasi-stationary sea surface topography is characterized by an increase from the polar regions toward the equator and by strong inclinations along the main ocean currents.

3.4.3   The Geoid as Height Reference Surface

The geoid is used in geodesy, cartography, and oceanography as a reference surface for heights and depths (continental and ocean bottom topography, as well as sea surface topography). A point P can be attributed to a specified level surface by its gravity potential W (Fig. 3.12). With respect to the geoid potential W_0, the "height" of P is given by the negative potential difference to the geoid, which is called the *geopotential number C*. We get from (3.52)

$$C = W_0 - W_P = -\int_{P_0}^{P} dW = \int g \, dn .$$ (3.104)

The integral is independent of the path; hence, P_0 is an arbitrary point on the geoid. C can be determined from geometric leveling and gravity measurements along any path between P_0 and P.
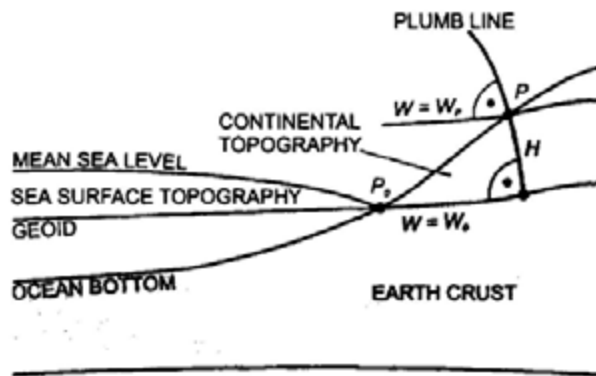
Fig. 3.12. Geoid, mean sea level, continental and sea surface topography

The geopotential number is an ideal measure for describing the behavior of masses (e.g., water masses) in the gravity field. It could be used as a "height" in several applications, as in hydraulic engineering and oceanography. A more general use is limited by the potential unit $m^2 s^{-2}$, which is in contradiction to the obvious demand for a metric height system that employs the "meter" unit.

In order to achieve a certain agreement with the numerical value of the height in meters, the geopotential unit (gpu) $10 \ m^2 s^{-2}$, or kgalm, is also used for the geopotential number. With $g \approx 9.8 \ ms^{-2}$, the values of C are about 2% smaller than the corresponding height values.

The *dynamic height* H^{dyn} is obtained by dividing the geopotential number through a constant gravity value. Usually the normal gravity γ_0^{45} calculated for the surface of the level ellipsoid at 45° latitude is used, cf. [4.2.2]:

$$H^{dyn} = \frac{C}{\gamma_0^{45}} .$$

(3.105)

The surfaces H^{dyn} = const. remain equilibrium surfaces. Hence, points located on the same level surface have the same dynamic height. Unfortunately, a geometric interpretation of the dynamic heights is not possible, and larger corrections are necessary in order to convert leveling results into dynamic height differences, cf. [6.4.1]. Because of this, dynamic heights have not been widely used in geodesy but are used in oceanography, cf. [3.4.2].

National or continental height systems, and terrain-data based on them (topographic maps, digital terrain models), use either orthometric or normal heights.

The *orthometric height H* is defined as the linear distance between the surface point and the geoid, reckoned along the curved plumb line (Fig. 3.12). This definition corresponds to the common understanding of "heights above sea level". Expanding the right-hand side of (3.104) in H and integrating along the plumb line from P_0 ($H = 0$) to P (H) we obtain

$$H = \frac{C}{\bar{g}}, \quad \bar{g} = \frac{1}{H} \int_0^H g \, dH \, . \tag{3.106}$$

$\bar{g}$ is the mean gravity along the plumb line; gravity values inside the earth are required for its calculation. This is performed by introducing a model of the density distribution of the topographic masses. As this distribution is known only imperfectly, the accuracy of computed orthometric heights depends on the accuracy of the density model. In addition, points of equal orthometric height deviate slightly from a level surface, which is due to the non-parallelism of the level surfaces, cf. [3.2.1]. These drawbacks are compensated by the fact that orthometric heights represent the geometry of the topographic masses. Geometric leveling only needs small corrections for the transformation into orthometric height differences, cf. [6.4.1].

In order to avoid any hypothesis on the distribution of the topographic masses, *normal heights H^N* have been introduced and are used in a number of countries. The mean gravity $\bar{g}$ in (3.106) is replaced by the mean normal gravity $\bar{\gamma}$ along the normal plumb line, which is only slightly curved:

$$H^N = \frac{C}{\bar{\gamma}}, \quad \bar{\gamma} = \frac{1}{H^N} \int_0^{H^N} \gamma \, dH^N \, . \tag{3.107}$$

$\bar{\gamma}$ can be calculated in the normal gravity field of an ellipsoidal earth model. The reference surface for the normal heights is the quasi-geoid, which is close to the geoid but not a level surface. It deviates from the geoid on the mm to cm-order at low elevations and may reach one-meter deviation in the high mountains. On the oceans, geoid and quasigeoid practically coincide, cf. [6.1.1].

The *zero height surface* (vertical datum) of national height systems generally is defined by the mean sea level derived from tide gauge records over a certain time interval. These reference surfaces only approximate the geoid, due to the sea surface topography and local anomalies, with deviations up to one meter, cf. [3.4.2].

The adjustment of continental-wide leveling networks provides heights that refer to one common level surface. By comparing with the mean sea level obtained at tide gauges, sea level slopes that partially agree with oceanic leveling results have been found, cf. [3.4.2]. MSL at the Pacific coast of the U.S.A. is about 1 m higher than at the Atlantic coast (ZILKOSKI et al. 1995), and the mean Baltic Sea level is estimated to be about 0.5 m above MSL of the Mediterranean Sea (GRONWALD 1963). On the other hand, there are also larger discrepancies of a few 0.1 m between the results of geometric and oceanic leveling. These can be traced back to differently defined reference surfaces, to the particular behavior of MSL along the coastlines, and to systematic errors in geometric leveling over long distances (FISCHER 1977). It should be mentioned that older leveling networks have often been adjusted without any gravity reduction, or by substituting actual gravity with normal gravity. A unification of the different vertical datum systems to a world-wide standard thus would require vertical shifts to a common reference (global geoid definition) and a uniform treatment of the height measurements (RUMMEL and TEUNISSEN 1988, RAPP 1995b), cf. [7.2].

3.5   Temporal Gravity Variations

Gravity changes with time may be divided into effects due to a time dependent gravitational constant and variations of the earth's rotation [3.5.1], tidal accelerations [3.5.2], and variations caused by terrestrial mass displacements [3.5.3]. These changes are of global, regional, or local character and occur either at well-known frequencies (tides) or at time scales ranging from secular to abrupt (LAMBECK 1988, MUELLER and ZERBINI 1989).

3.5.1   Gravitational Constant, Earth Rotation

Based on cosmological considerations, DIRAC (1938) postulated a secular decrease of the *gravitational constant* G, with relative changes of $\dot{G}/G = -10^{-10}$ to $-10^{-11}/a$ $\left(\dot{G} = dG/dt\right)$. Even to this day, laboratory experiments and the analysis of long-term observations to artificial satellites and the moon have not supported the assumption $\dot{G} \neq 0$ (GILLIES 1987, BURŠA 1995).

The *earth's rotational vector* $\boldsymbol{\omega}$ is subject to secular, periodic, and irregular variations, leading to changes of the centrifugal acceleration $\mathbf{z}$, cf. [2.5.2]. In a spherical approximation, the radial component of $\mathbf{z}$ enters into gravity, cf. [3.1.4]. By multiplying (3.35b) with $\cos\overline{\varphi}$ ($\overline{\varphi}$ = geocentric latitude), we obtain

$$z_r = -\omega^2 r \cos^2 \overline{\varphi} \, . \tag{3.108}$$

Differentiation yields the effect of changes in latitude (polar motion) and angular velocity (length of day) on gravity:

$$\delta z_r = \omega^2 r \sin 2\overline{\varphi}\, d\overline{\varphi} - 2\omega r \cos^2 \overline{\varphi}\, d\omega . \tag{3.109}$$

Polar motion does not exceed a few 0.1"/a, and rotation changes are at the order of a few ms. Hence, corresponding gravity variations on the earth's surface ($r = 6371$ km) remain less than $0.1\mu\,\mathrm{m\,s}^{-2}$ and $0.01\mu\,\mathrm{m\,s}^{-2}$ respectively.

3.5.2   Tidal Acceleration, Tidal Potential

Tidal acceleration is caused by the superposition of lunisolar gravitation (and to a far lesser extent planetary gravitation) and orbital accelerations due to the motion of the earth around the barycenter of the respective two-body system (earth-moon, earth-sun etc.). The periods of these orbital motions are about 28 days for the moon and 365 days for the sun, and the gravimetric tidal effect is on the order of $10^{-7}\,g$ (MELCHIOR 1983, ZUERN and WILHELM 1984, WENZEL 1997a).

For a *rigid earth*, the tidal acceleration at a given point can be determined from Newton's law of gravitation and the ephemerides (coordinates) of the celestial bodies (moon, sun, planets). The computations are carried out separately for the individual two-body systems (earth-moon, earth-sun etc.), and the results are subsequently added, with the celestial bodies regarded as point masses.

We consider the geocentric coordinate system to be moving in space with the earth but not rotating with it (revolution without rotation). All points on the earth experience the same orbital acceleration in the geocentric coordinate system (see Fig. 3.13 for the earth-moon system). In order to obtain equilibrium, orbital acceleration and gravitation of the celestial bodies have to cancel in the earth's center of gravity. Tidal acceleration occurs at all other points of the earth. The acceleration is defined as the difference between the gravitation $\mathbf{b}$, which depends on the position of the point, and the constant part $\mathbf{b}_0$, referring to the earth's center:

$$\mathbf{b}_t = \mathbf{b} - \mathbf{b}_0 . \tag{3.110}$$

The tidal acceleration deforms the earth's gravity field symmetrically with respect to three orthogonal axes with origin at the earth's center. This tidal acceleration field experiences diurnal and semidiurnal variations, which are due to the rotation of the earth about its axis.

If we apply the law of gravitation to (3.110), we obtain for the moon (m)
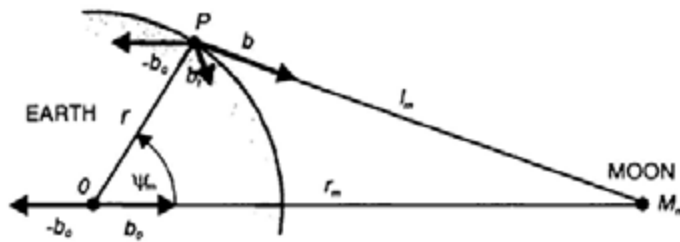
Fig. 3.13. Lunar gravitation, orbital acceleration, and tidal acceleration

$$\mathbf{b}_t = \frac{GM_m}{l_m^2}\frac{\mathbf{l}_m}{l_m} - \frac{GM_m}{r_m^2}\frac{\mathbf{r}_m}{r_m}.\tag{3.111}$$

Here, M_m = mass of the moon, and l_m and r_m = distance to the moon as reckoned from the calculation point P and the earth's center of gravity O respectively. We have $\mathbf{b}_t = 0$ for $l_m = r_m$. Corresponding relations hold for the earth-sun and earth-planet systems.

We now make the transition from the tidal acceleration to the *tidal potential*:

$$\mathbf{b}_t = \operatorname{grad} V_t = \operatorname{grad}(V - V_0).\tag{3.112}$$

In the geocentric system, using spherical coordinates r_m, ψ_m, the law of gravitation yields the potential of a point mass according to (3.9):

$$V = \frac{GM_m}{l_m}\tag{3.113a}$$

with

$$l_m = \left(r^2 + r_m^2 - 2rr_m \cos\psi_m\right)^{\frac{1}{2}}.\tag{3.113b}$$

The potential of the homogeneous $\mathbf{b}_0$-field is given by multiplying b_0 with $r \cos\psi_m$:

$$V_0 = \frac{GM_m}{r_m^2} r \cos\psi_m.\tag{3.114}$$

Inserting (3.113) and (3.114) into (3.112), and adding an integration constant so that $V_t = 0$ for $r = 0$ and $l_m = r_m$, we get for the tidal potential

$$V_t = GM_m \left(\frac{1}{l_m} - \frac{1}{r_m} - \frac{r\cos\psi_m}{r_m^2} \right). \tag{3.115}$$

The tidal potential, and functionals thereof, can be calculated either from the ephemerides of the celestial bodies or from a spherical harmonic expansion. Tidal potential catalogues are based primarily on the latter method, as the series expansions converge rapidly close to the earth's surface ($r = R$), with $r/r_m = 1/60$ for the moon and a corresponding relation of $1/23600$ for the sun. The results from calculations employing the ephemerides serve as a control for the tidal potential catalogues.

We develop (3.113b) into a series according to (3.79). When inserting into (3.115), the terms of degree zero and one cancel and we obtain

$$V_t = \frac{GM_m}{r_m} \sum_{l=2}^{\infty} \left(\frac{r}{r_m} \right)^l P_l(\cos\psi_m), \tag{3.116}$$

where $P_l(\cos\psi_m)$ are the Legendre polynomials. The largest contribution ($\approx 98\%$) stems from degree two. Restricting ourselves now to the degree $l = 2$, and inserting P_2 (3.83) in the form

$$\cos^2\psi_m = \frac{1}{2}(\cos 2\psi_m + 1),$$

we get the *main term* of the tidal potential series

$$V_t = \frac{3}{4}GM_m \frac{r^2}{r_m^3} \left(\cos 2\psi_m + \frac{1}{3} \right). \tag{3.117}$$

For $r = R$, and neglecting the slight variation of r_m, the expression before the parentheses is called *Doodson's* tidal constant. It is $2.628\,\mathrm{m^2\,s^{-2}}$ for the moon and $1.208\,\mathrm{m^2\,s^{-2}}$ for the sun. Hence, the solar tides amount to 46% of the lunar tides.

Differentiating (3.117) generates the *tidal acceleration*. The *radial* component (positive outward) is found to be

$$b_r = \frac{\partial V_t}{\partial r} = \frac{3}{2}GM_m \frac{r}{r_m^3} \left(\cos 2\psi_m + \frac{1}{3} \right). \tag{3.118}$$

The *tangential* component (positive in the direction toward the moon) is

$$b_\psi = -\frac{\partial V_t}{r\partial\psi} = \frac{3}{2}GM_m\frac{r}{r_m^3}\sin 2\psi_m.$$ (3.119)

Equations (3.117) to (3.119) permit calculation of the tidal effects on the level surfaces, on gravity, and on the plumb line direction for a rigid earth.

Taking the relation (3.52) between a potential change and the vertical shift of a level surface into account, we obtain a tidal-induced increase of the level surface. This amounts to $0.36\,\text{m}$ for the moon and $0.16\,\text{m}$ for the sun at $\psi = 0°$ and $180°$ respectively. At $\psi = 90°$ and $270°$, we have a decrease of $0.18\,\text{m}$ and $0.08\,\text{m}$ respectively. For stationary systems, the level surfaces would experience a corresponding deformation, and freely moving masses of water would assume the form of one of these surfaces (equilibrium tide), Fig. 3.14.
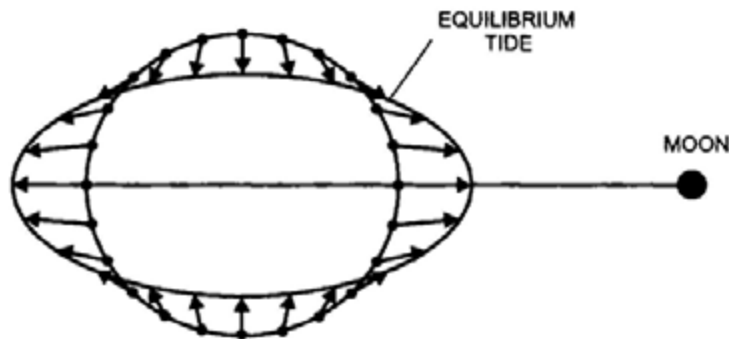
Fig. 3.14. Tidal acceleration and equilibrium tide

According to (3.118), *gravity* changes (opposite sign) would vary between $-1.1\mu\text{ms}^{-2}$ (moon) and $-0.5\mu\text{ms}^{-2}$ (sun) for $\psi = 0°$ (zenithal position); and $+0.5\mu\text{ms}^{-2}$ (moon) and $+0.3\mu\text{ms}^{-2}$ (sun) for $\psi = 90°$ and $135°$. Changes in the direction of the *plumb line* are given by b_ψ/g. There is no tidal effect at $\psi = 0°$ and $90°$. Maximum values occur at $\psi = 45°$ and $135°$, with fluctuations of $\pm 0.017''$ (moon) and $\pm 0.008''$ (sun).

Equation (3.117) provides the dependence of the tidal potential on the zenith angle (and the distance) to the celestial body. The variation of the tidal potential and acceleration is more easily recognized if we change to the earth-fixed coordinate system $(\overline{\varphi},\lambda)$ for the point of calculation and to the equatorial system of astronomy (δ,h) for the celestial body, cf. [2.4.1]. Following (2.21), we have (for the moon) the relation

$$\cos\psi_m = \sin\overline{\varphi}\sin\delta_m + \cos\overline{\varphi}\cos\delta_m\cos h_m \qquad (3.120)$$

with the hour angle given by (2.22) and (2.23):

$$h_m = \text{LAST} - \alpha_m = \lambda + \text{GAST} - \alpha_m. \qquad (3.121)$$

Substituting into (3.117) yields *Laplace's tidal equation* (for the moon):

$$V_t = \frac{3}{4}GM_m\frac{r^2}{r_m^3}\left\{\begin{array}{l}\left(\dfrac{1}{3}-\sin^2\overline{\varphi}\right)\left(1-3\sin^2\delta_m\right)+\\[2mm]\sin 2\overline{\varphi}\sin 2\delta_m\cos h_m + \cos^2\overline{\varphi}\cos^2\delta_m\cos 2h_m\end{array}\right\}. \qquad (3.122)$$

The quantities r_m, δ_m, and h_m vary with time, having different periods. The first term, which is independent of the earth's rotation, exhibits long-periodic variations (14 days for the moon, 1/2 year for the sun). It also contains a non-periodic part, which only depends on latitude, causing a permanent deformation of the level surfaces including the geoid, cf. [3.4.1]. Using (3.52), and taking the inclination of the ecliptic into account, the geoid is thus lowered by 0.19 m at the poles and raised by 0.10 m at the equator (EKMAN 1989).

The second term oscillates with diurnal periods because of the daily rotation of the earth as expressed by the hour angle h, and the third term introduces semidiurnal periods. Long-periodic terms enter through the declination δ and the right ascension α. As seen from (3.122), long-periodic and semidiurnal tides are symmetric about the equator, while the diurnal tides are antisymmetric. The diurnal tide has its maximum at $\varphi = \pm 45°$ and vanishes at the equator and the poles, while the semidiurnal tide reaches its maximum at the equator and is zero at the poles. The long-periodic tides have a maximum at the poles.

Each of the three tidal constituents in (3.122) varies in a complicated way, since they contain products of different time varying functions. However, the ephemerides of the moon and the sun can be expressed as harmonic functions of five fundamental astronomic quantities, considering that these quantities essentially change uniformly with time (MELCHIOR 1983). Introducing these harmonic series into (3.122) yields a spectral analysis of the tidal potential, and with (3.118) and (3.119) we get a spectral analysis of the tidal acceleration. Thus, potential and acceleration are represented by the sum of time-dependent cosine functions having constant periods and amplitudes and phases that depend on latitude and height (partial tides). Tab. 3.1 gives the periods and amplitudes of the main gravimetric partial tides for $\varphi = 45°$.

A first expansion for the moon and the sun was carried out by DOODSON (1921). The development by CARTWRIGHT and TAYLER (1971) and CARTWRIGHT and EDDEN (1973) contains 505 partial tides (uncertainty less than $1\,\text{nms}^{-2}$) and was recommended by IAG for the computation of the tides of the rigid earth (RAPP 1983b). A recent tidal catalogue by HARTMANN and WENZEL (1995)

is based on a spherical harmonic development to degree 6 (moon) and degree 3 (sun) and includes the effects of Venus, Mars, and Jupiter (4 orders of magnitude smaller than the tidal effects of moon and sun). It also takes the flattening of the earth into account. This catalogue provides 12 935 partial tides with an accuracy of $\pm 0.001 \, \text{nm} \, \text{s}^{-2}$, WENZEL (1996).

As the earth is not a rigid body, it reacts in a different way to the tidal force. The solid earth behaves mainly as an elastic body: *earth's body tides*. In the oceans, tidal oscillations depend on the ocean-bottom topography, with large differences occurring at the coastlines and at the shelf areas: *ocean tides*, ZAHEL (1997). While the measurement of tidal effects will be discussed in [5.4.6] and [5.5.4], the theory of earth tides and results of earth-tide observations are given in [8.3.5].

Tab. 3.1. Principal Gravimetric Partial Tides for $\bar{\varphi} = 45°$, $h = 0$

symbol	name	period (solar hours)	amplitude $\left(\text{nm} \, \text{s}^{-2} \right)$
long-periodic waves			
M0	const. *l* tide	∞	102.9
S0	const. *s* tide	∞	47.7
Ssa	declin. tide to S0	182.62 d	14.8
Mm	ellipt. tide to M0	27.55 d	16.8
Mf	declin. tide to M0	13.66 d	31.9
diurnal waves			
O1	main diurnal *l* tide	25.82 h	310.6
P1	main diurnal *s* tide	24.07 h	144.6
Q1	ellipt. tide to O1	26.87 h	59.5
K1	main diurnal *ls* decl. tide	23.93 h	436.9
semi-diurnal waves			
M2	main *l* tide	12.42 h	375.6
S2	main *s* tide	12.00 h	174.8
N2	ellipt. tide to M2	12.66 h	71.9
K2	declin. tide to M2, S2	11.97 h	47.5
ter-diurnal waves			
M3	terdiurn. *l* tide	8.28 h	5.2

3.5.3   Non-tidal Temporal Variations

The terrestrial gravity field is affected by a number of variations with time due to mass redistributions in the atmosphere, the hydrosphere, and the solid earth. These processes take place at different time scales and are of global, regional, and local character (WAGNER and McADOO 1986, MUELLER and ZERBINI 1989, TORGE 1993).

Long-term global effects include postglacial rebound, melting of the ice caps and glaciers, as well as sea level changes induced by atmospheric warming; slow motions of the earth's core and mantle convection also contribute. Subsidence in sedimentary basins and tectonic uplift are examples of regional effects. Groundwater variations are primarily of seasonal character, while volcanic and earthquake activities are short-term processes of more local extent.

The magnitude of the resulting gravity variations depends on the amount of mass shifts and is related to them by the law of gravitation. As seen from (3.5), a shift of the observer also affects gravitation, especially where height changes play a dominant role. Research and modeling of these variations is still in the beginning stages. Large-scale variations have been found from satellite-derived gravity field models, but small-scale effects can be detected only by terrestrial gravity measurements. Simple models have been developed for the relation between atmospheric and hydrological mass shifts and gravity changes, cf. [5.4.1]. Generally, gravity changes produced by mass redistributions do not exceed the order of 10^{-9} to $10^{-8} g$, while geoid changes remain less than 1 mm/a, cf. [8.3.4].

4 The Geodetic Earth Model

A geodetic earth model is used as a reference for the actual surface and external gravity field of the earth. It should provide a good fit to the geoid and to the gravity field, and thus allow the linearization of non-linear geodetic problems. On the other hand, the mathematical formation of the model should be simple and possibly permit calculations by closed formulas. The model should serve as a standard for applications not only in geodesy and cartography but also in astronomy and geophysics; it should satisfy the demands and needs of these disciplines too.

Based on these considerations, the level ellipsoid has been introduced. It possesses a simple geometry, and coordinate systems that refer to it approximate the "natural" coordinate system sufficiently well [4.1]. The ellipsoid's mass and rotation provide a "normal" gravity field, which can be calculated if the ellipsoid surface is defined to be in equilibrium [4.2]. State of the art earth models are recommended from time to time as a standard and are given the name Geodetic Reference System [4.1.3].

4.1 The Rotational Ellipsoid

The rotational ellipsoid was introduced as a geometrical figure of the earth in the 18th century, cf. [1.3.2]. By fitting its dimension and orientation to the geoid, it approximates this level surface within about 100 m. The geometry of the ellipsoid can be described in a simple manner, together with ellipsoidal surface coordinates and curvature [4.1.1], [4.1.2]. The use of global and local three-dimensional ellipsoidal systems provides an approximation to the corresponding systems of the actual earth and permits the separation between horizontal position and height [4.1.3].

Geometry and coordinate systems of the ellipsoid are well documented in geodetic literature, e.g., GROSSMANN (1976), BOMFORD (1980), HECK (1995), HEITZ (1988).

4.1.1 Parameters and Coordinate Systems

The rotational ellipsoid is generated by rotating the meridian ellipse about its minor axis. The shape of the ellipsoid is described by two geometric parameters, the *semimajor axis* a and the *semiminor axis* b (Fig. 4.1). Generally, b is replaced by a smaller quantity, describing the (small) polar flattening of the
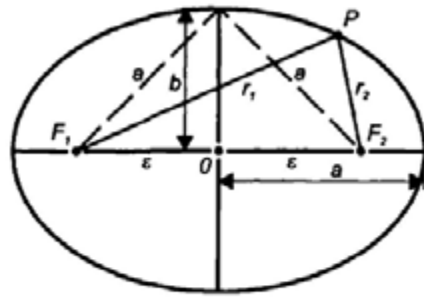
Fig. 4.1. Meridian Ellipse

ellipsoid, which is more suitable for series expansions. We have the (geometrical) *flattening*

$$f = \frac{a-b}{a},\tag{4.1a}$$

the *first eccentricity*

$$e = \frac{\sqrt{a^2 - b^2}}{a},\tag{4.1b}$$

and the *second eccentricity*

$$e' = \frac{\sqrt{a^2 - b^2}}{b}.\tag{4.1c}$$

The following relations hold among those quantities:

$$\frac{b}{a} = 1 - f = \sqrt{1 - e^2} = \frac{1}{\sqrt{1 + e'^2}} = \frac{e}{e'}.\tag{4.2}$$

From the geometric definition of the ellipse as the curve having a constant value for the sum of the distances r_1 and r_2 to the focal points F:

$$r_1 + r_2 = 2a,$$

we derive the *linear eccentricity*

$$\varepsilon = \sqrt{a^2 - b^2}.\tag{4.3}$$

Fig. 4.2. Geodetic coordinates latitude and longitude

We now introduce a spatial $\overline{X}, \overline{Y}, \overline{Z}$-Cartesian coordinate system (Fig. 4.2). The origin of the system is situated at the center of the ellipsoid O. The $\overline{Z}$-axis coincides with the minor axis of the ellipsoid. The equation of the surface of the ellipsoid is then given by

$$\frac{\overline{X}^2 + \overline{Y}^2}{a^2} + \frac{\overline{Z}^2}{b^2} - 1 = 0 . \tag{4.4}$$

The system of *geodetic surface coordinates* is defined by the ellipsoidal *latitude* φ and *longitude* λ (also geodetic latitude and longitude). φ is the angle measured in the meridian plane between the equatorial plane ($\overline{X}, \overline{Y}$-plane) of the ellipsoid and the surface normal at P. Longitude λ is the angle measured in the equatorial plane between the zero meridian ($\overline{X}$-axis) and the meridian plane of P. Here, φ is positive northward and negative southward, and λ is positive as reckoned toward the east. The ellipsoidal *meridian plane* is formed by the surface normal and the $\overline{Z}$-axis. φ and λ are defined to have angular values, but they may also be considered as curvilinear surface coordinates. The *coordinate lines* of this orthogonal system are the meridians ($\lambda = $ const.) and the parallels, or circles of latitude, ($\varphi = $ const.). With

$$\overline{X} = p\cos\lambda, \quad \overline{Y} = p\sin\lambda , \tag{4.5}$$

we introduce the radius of the circle of latitude

$$p = \sqrt{\overline{X}^2 + \overline{Y}^2} \tag{4.6}$$

Fig. 4.3. Geodetic, reduced, and geocentric latitude

as a new variable (Fig. 4.2). Substituting p into (4.4) and differentiating yields the slope of the ellipsoidal tangent at P (Fig. 4.3):

$$\frac{d\bar{Z}}{dp} = -\left(\frac{b}{a}\right)^2 \frac{p}{\bar{Z}} = -\cot\varphi. \qquad (4.7)$$

By combining (4.4) and (4.7), and substituting p with (4.5), the parametric representation of the *meridian ellipse* follows:

$$\bar{X} = \frac{a^2 \cos\varphi \cos\lambda}{\left(a^2 \cos^2\varphi + b^2 \sin^2\varphi\right)^{1/2}}, \quad \bar{Y} = \frac{a^2 \cos\varphi \sin\lambda}{\left(a^2 \cos^2\varphi + b^2 \sin^2\varphi\right)^{1/2}},$$

$$\bar{Z} = \frac{b^2 \sin\varphi}{\left(a^2 \cos^2\varphi + b^2 \sin^2\varphi\right)^{1/2}}. \qquad (4.8)$$

Instead of φ, other latitude parameters are used for special applications. The *geocentric latitude* $\bar{\varphi}$ has already been introduced together with the longitude λ and the geocentric distance r as spherical coordinate, cf. [2.5.1]. From Fig. 4.3, the corresponding equation of the *ellipse* is given by

$$p = r\cos\bar{\varphi}, \quad \bar{Z} = r\sin\bar{\varphi}, \qquad (4.9)$$

where p follows from (4.5).

The *reduced latitude* β is obtained by projecting (parallel to the $\bar{Z}$-axis) from the ellipse to the concentric circle of radius a (Fig. 4.3). Since the ratio of the elliptical to the circular ordinates is b/a (ellipse as the affine image of the circle), we have

$$p = a\cos\beta, \ \overline{Z} = \frac{b}{a}a\sin\beta = b\sin\beta. \tag{4.10}$$

Using β instead of φ formally transforms ellipsoidal formulas into spherical formulas.

Comparing (4.9) and (4.10) with (4.7) provides the transformation between φ, $\overline{\varphi}$, and β:

$$\tan\overline{\varphi} = \left(\frac{b}{a}\right)^2 \tan\varphi = \left(1 - e^2\right)\tan\varphi$$
$$\tan\beta = \frac{b}{a}\tan\varphi = \sqrt{1 - e^2} \ \tan\varphi \tag{4.11a}$$

A series expansion yields the differences in the angles:

$$\varphi - \overline{\varphi} = \frac{e^2}{2}\sin 2\varphi + \cdots = 2\left(\varphi - \beta\right). \tag{4.11b}$$

The maximum difference occurs at $\varphi = 45°$, with $\left(\varphi - \overline{\varphi}\right) = 690''$.

4.1.2 Curvature

The meridians and parallels are the *lines of curvature* of the rotational ellipsoid. The principal *radii of curvature* are therefore in the plane of the meridian and in the plane of the prime vertical perpendicular to the meridian plane (Fig. 4.4).

The curvature of the *meridian* (M = curvature radius) as a plane curve $\overline{Z} = \overline{Z}(p)$ in the $\overline{Z}, p$-plane is given by

$$\frac{1}{M} = -\frac{d^2\overline{Z}/dp^2}{\left(1 + \left(d\overline{Z}/dp\right)^2\right)^{3/2}}. \tag{4.12}$$

With (4.7) and its derivative, and taking (4.2) into account, we obtain the meridian radius of curvature

$$M = \frac{a\left(1 - e^2\right)}{\left(1 - e^2\sin^2\varphi\right)^{3/2}}. \tag{4.13}$$
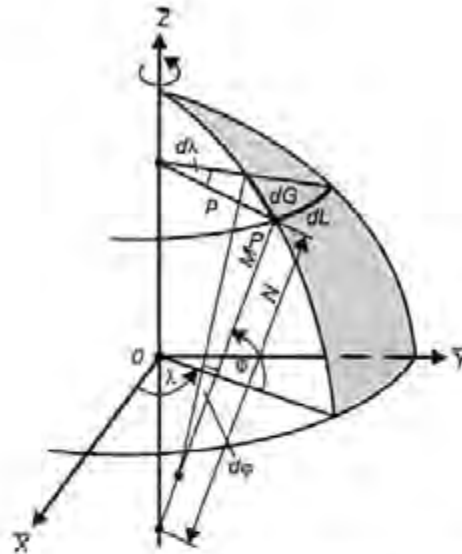
Fig. 4.4. Curvature of the rotational ellipsoid

The plane of a parallel circle (oblique section of the rotational ellipsoid) and the vertical plane in the same tangential direction intersect at point P with the angle φ. The theorem of *Meusnier* (regarding surface curvatures, see, e.g., STOKER 1969) provides the radius of curvature in the *prime vertical* :

$$N = \frac{p}{\cos\varphi}.$$
(4.14)

Because of rotational symmetry, the origin of N is on the spin axis. Inserting (4.6) and (4.8) into (4.14), one obtains

$$N = \frac{a}{\left(1 - e^2 \sin^2\varphi\right)^{1/2}}.$$
(4.15)

A comparison of (4.13) and (4.15) shows that $N \geq M$. At the poles $(\varphi = \pm 90°)$, the polar radius of curvature becomes

$$c = M_{90} = N_{90} = \frac{a^2}{b}.$$
(4.16)

At the equator $(\varphi = 0°)$ the values are

$$M_0 = \frac{b^2}{a}, \quad N_0 = a. \tag{4.17}$$

The curvature of the *normal section* of an ellipsoidal, or geodetic, azimuth α is computed according to *Euler's* formula by

$$\frac{1}{R_\alpha} = \frac{\cos^2 \alpha}{M} + \frac{\sin^2 \alpha}{N}. \tag{4.18}$$

Here, R_α is the radius of curvature. The *geodetic azimuth* α is defined as the angle measured in the horizontal plane between the ellipsoidal meridian plane of P and the vertical plane containing the normal to P and the target point; α is reckoned from north in the clockwise direction. The *mean curvature* $\bar{J}$ is given by

$$\bar{J} = \frac{1}{2}\left(\frac{1}{M} + \frac{1}{N}\right). \tag{4.19}$$

The *arc lengths* of the *coordinate lines* of the φ,λ-system are computed using M and N. For the arc elements of the meridian and the parallel respectively, we obtain (Fig. 4.4)

$$dG = M\, d\varphi, \quad dL = N \cos\varphi\, d\lambda. \tag{4.20}$$

With (4.13), the length of the *meridian arc* (starting at the equator) becomes

$$G = \int_0^\varphi M\, d\varphi = a(1-e^2) \int_0^\varphi \frac{d\varphi}{(1-e^2 \sin^2 \varphi)^{3/2}}. \tag{4.21a}$$

Equation (4.21a) can be reduced to an elliptic integral of the second kind, which cannot be evaluated in a closed form (KUTTERER 1998). Practical computations may be based on numerical integration (e.g., by Simpson's rule) or on a binomial expansion of the denominator. Subsequent term-by-term integration yields

$$G = a(1-e^2)\left(\left(1 + \frac{3}{4}e^2 + \cdots\right)\varphi - \left(\frac{3}{8}e^2 + \cdots\right)\sin 2\varphi + \cdots\right). \tag{4.21b}$$

Short arcs $(\Delta\varphi = \varphi_2 - \varphi_1 < 1°)$ can be calculated by a rapidly converging Taylor expansion. Expanding about the middle latitude $\varphi_M = (\varphi_1 + \varphi_2)/2$ yields

$$\Delta G_{1,2} = G_2 - G_1 = \left(\frac{dG}{d\varphi}\right)_{\varphi_M} \Delta\varphi + \cdots. \tag{4.21c}$$

According to (4.20), the arc length of a *circle of latitude* between the geodetic longitudes λ_1 and λ_2 is given by

$$\Delta L = \int_{\lambda_1}^{\lambda_2} N\cos\varphi\, d\lambda = N\cos\varphi\left(\lambda_2 - \lambda_1\right). \tag{4.22}$$

With $a = 6\,378\,137$ m, $b = 6\,356\,752$ m, and $e^2 = 0.006\,694\,380$ (for numerical values see [4.3]), we get for the radii of curvature at the poles and at the equator

$$c = 6\,399\,594 \text{ m}, \quad M_0 = 6\,335\,439 \text{ m}, \quad N_0 = a.$$

The arc lengths along the meridian and the parallel for $\varphi = 50°$ are

$$\Delta G\left(\Delta\varphi = 1°\right) = 111\,229\,\text{m}, \qquad \Delta L\left(\Delta\lambda = 1°\right) = 71\,696\,\text{m},$$

$$\Delta G\left(\Delta\varphi = 1'\right) = 1853.8\,\text{m}, \qquad \Delta L\left(\Delta\lambda = 1'\right) = 1194.9\,\text{m},$$

$$\Delta G\left(\Delta\varphi = 1''\right) = 30.90\,\text{m}, \qquad \Delta L\left(\Delta\lambda = 1''\right) = 19.92\,\text{m}.$$

Local approximations to the ellipsoid use the *Gaussian osculating sphere* of radius

$$R_G = \sqrt{M\left(\varphi_0\right)N\left(\varphi_0\right)}. \tag{4.23}$$

At the latitude φ_0 it has the same Gaussian curvature as the ellipsoid.

Global approximations can be based on a sphere with the *mean radius*

$$R_m = \frac{1}{3}\left(2a + b\right), \tag{4.24a}$$

the radius derived from equality of *volumes* (i.e., volume of sphere = volume of ellipsoid)

$$R_V = \sqrt[3]{a^2 b}, \tag{4.24b}$$

or the radius for a sphere having a surface area equal to that of the ellipsoid. The latter one results from an integration over the ellipsoidal surface elements dG and dL (4.20), which after a series expansion yields

$$R_s = b\left(1 + \frac{2}{3}e^2 + \frac{3}{5}e^4 + \cdots\right)^{1/2}. \tag{4.24c}$$

The numerical values for these three approaches agree within a few meters, which leads to a global value of $R = 6371$ km.

4.1.3 Spatial Geodetic Coordinates

The ellipsoidal surface coordinate system (φ,λ) can be extended to a spatial system by introducing the height h of the point P above the ellipsoid measured along the surface normal (Fig. 4.5). The point Q on the ellipsoid thus is obtained by projecting the point P along the ellipsoidal normal (Helmert projection). The spatial coordinates φ,λ,h are designated as geodetic coordinates.
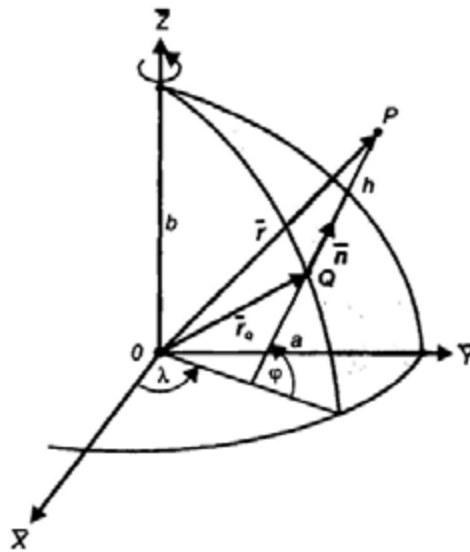
Fig. 4.5. Spatial geodetic coordinates

The *coordinate surfaces* (φ = const., λ = const., h = const.) of this system are orthogonal. The *coordinate lines* (φ-line = geodetic meridian, λ-line = geodetic parallel, h-line = ellipsoidal normal) represent planar curves.

In (4.8) we substitute the first eccentricity e^2 for the semiminor axis b, taking (4.15) into account; the coordinate vector for the point Q on the ellipsoid (4.8) then transforms into

$$\bar{r}_Q = \begin{pmatrix} \bar{X}_Q \\ \bar{Y}_Q \\ \bar{Z}_Q \end{pmatrix} = N \begin{pmatrix} \cos\varphi\cos\lambda \\ \cos\varphi\sin\lambda \\ (1-e^2)\sin\varphi \end{pmatrix}. \qquad (4.25)$$

For the point P, we get according to Fig. 4.5

$$\bar{r} = \bar{r}_Q + h\bar{n}, \qquad (4.26a)$$

with the surface normal

$$\bar{n} = \begin{pmatrix} \cos\varphi\cos\lambda \\ \cos\varphi\sin\lambda \\ \sin\varphi \end{pmatrix}, \qquad (4.26b)$$

or

$$\bar{r} = \begin{pmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \end{pmatrix} = \begin{pmatrix} (N+h)\cos\varphi\cos\lambda \\ (N+h)\cos\varphi\sin\lambda \\ ((1-e^2)N+h)\sin\varphi \end{pmatrix}. \qquad (4.27)$$

The *inverse problem* can be solved for φ and h only by iterative methods. From (4.27) we get (HEISKANEN and MORITZ 1967, p.183)

$$h = \frac{\sqrt{\bar{X}^2+\bar{Y}^2}}{\cos\varphi} - N, \quad \varphi = \arctan\frac{\bar{Z}}{\sqrt{\bar{X}^2+\bar{Y}^2}}\left(1 - e^2\frac{N}{N+h}\right)^{-1}$$

$$\lambda = \arctan\frac{\bar{Y}}{\bar{X}} \qquad (4.28)$$

The iteration process may start with $h=0$, which results in a first approximation for φ, and so on. Close to the earth's surface $(h \ll N)$ the process converges quickly. Efficient methods have also been developed for large heights (BORKOWSKI 1989, SJÖBERG 1999). Closed formulas with negligible residual
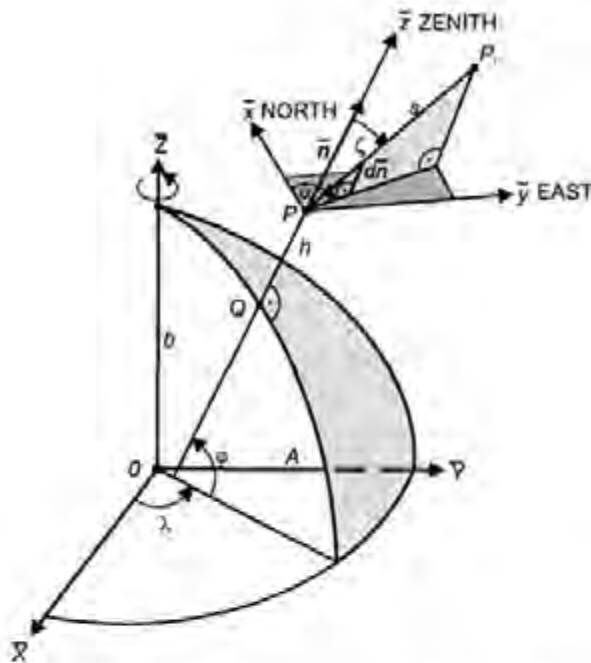
Fig. 4.6. Global and local ellipsoidal system

errors are given by BOWRING (1985). The transformation (4.28) is a standard problem in satellite geodesy, cf. [6.2.1].

Local ellipsoidal (geodetic) *systems* are introduced in analogy to the local astronomic systems, cf. [2.6.2], and represent an approximation to them (Fig. 4.6). With the origin at the point P, the local system is connected to the ellipsoidal vertical (outer surface normal $\bar{n}$ to the ellipsoid) through the geodetic latitude and longitude (4.26). The $\bar{z}$-axis is directed towards the ellipsoidal zenith, with the $\bar{x},\bar{y}$-plane being perpendicular to it. The $\bar{x}$-axis points to the ellipsoidal north (direction of the meridian), and the $\bar{y}$-axis points towards east (left-handed system).

A target point P_i is described with respect to P by the geodetic (ellipsoidal) *azimuth* α, introduced already in [4.1.2], the ellipsoidal *zenith angle* ζ, and the straight *distance* s between P and P_i. The zenith angle is measured in the vertical plane between the ellipsoidal vertical and the connecting line and reckoned positively from the zenith. These polar coordinates can be transformed into the local $\bar{x},\bar{y},\bar{z}$-system by a relation corresponding to (2.20):

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} = s \begin{pmatrix} \cos\alpha\sin\zeta \\ \sin\alpha\sin\zeta \\ \cos\zeta \end{pmatrix}. \tag{4.29}$$

After applying the reflection matrix $\mathbf{S}_2$ (2.24), the local system is transformed to the global $\bar{X}, \bar{Y}, \bar{Z}$-system by the rotation matrices $\mathbf{R}_2(90° - \varphi)$ and $\mathbf{R}_3(180° - \lambda)$, which correspond to (2.25) and (2.26):

$$\Delta\bar{\mathbf{X}} = \bar{\mathbf{A}}\bar{\mathbf{x}},\tag{4.30}$$

with

$$\Delta\bar{\mathbf{X}} = \left(\Delta\bar{X}, \quad \Delta\bar{Y}, \quad \Delta\bar{Z}\right)^T\tag{4.31}$$

and

$$\bar{\mathbf{A}} = \mathbf{R}_3\left(180° - \lambda\right)\mathbf{R}_2\left(90° - \varphi\right)\mathbf{S}_2$$
$$= \begin{pmatrix} -\sin\varphi\cos\lambda & -\sin\lambda & \cos\varphi\cos\lambda \\ -\sin\varphi\sin\lambda & \cos\lambda & \cos\varphi\sin\lambda \\ \cos\varphi & 0 & \sin\varphi \end{pmatrix}.\tag{4.32}$$

The inversion of (4.32) gives

$$\bar{\mathbf{x}} = \bar{\mathbf{A}}^{-1}\Delta\bar{\mathbf{X}},\tag{4.33}$$

with

$$\bar{\mathbf{A}}^{-1} = \bar{\mathbf{A}}^T = \begin{pmatrix} -\sin\varphi\cos\lambda & -\sin\varphi\sin\lambda & \cos\varphi \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\varphi\cos\lambda & \cos\varphi\sin\lambda & \sin\varphi \end{pmatrix},\tag{4.34}$$

which corresponds to (2.29) and (2.30).

4.2   The Normal Gravity Field

A "normal" gravity field may be referenced to the rotational ellipsoid by considering the latter to be a "level" ellipsoid. This earth model is now generally accepted as a geodetic reference system; higher order models do not offer any advantage [4.2.1]. The external gravity field of the level ellipsoid can be determined unambiguously from the parameters defining it [4.2.2]. The geometry of the normal gravity field is of special interest for geodetic applications [4.2.3].

4.2.1   The Level Ellipsoid, Level Spheroids

By adding the total mass M and the rotational angular velocity ω to the geometric parameters a and f of the rotational ellipsoid, we introduce an ellipsoidal gravity field composed of gravitation and centrifugal acceleration: *normal gravity field*. In addition, we require the surface of this ellipsoid to be a level surface of its own gravity field. According to the theorem of *Stokes-Poincaré*, the gravity field then is uniquely defined in the space exterior to the ellipsoid.

Theorem of *Stokes-Poincaré*: if a body of total mass M rotates with constant angular velocity ω about a fixed axis, and if S is a level surface of its gravity field enclosing the entire mass, then the gravity potential in the exterior space of S is uniquely determined by M, ω, and the parameters defining S.

The earth model thus defined is called a *level* (or equipotential) *ellipsoid*. Instead of a, f, M and ω, other sets of four independent parameters may be used for its definition. If the parameters are given values which correspond to the real earth, then an optimum approximation to the geometry of the geoid and to the external gravity field is achieved: *mean earth ellipsoid*, cf. [6.8.1]. The theory of the level ellipsoid has been developed by *P. Pizetti* (1894), *C. Somigliana* (1929), and others (HEISKANEN and MORITZ 1967, p.64).

From the physical point of view, an earth model is required which is in *hydrostatic equilibrium*. All its level surfaces then coincide with the surfaces of equal density and equal pressure. Deviations from this model would indicate stress in the earth's body, cf. [8.1]. The theory of equilibrium figures has been discussed since the days of *Newton* and *Clairaut* , cf. [1.3.2], see LEDERSTEGER (1956/1969), MORITZ (1990).

In the above definition of the level ellipsoid, nothing has been stated regarding the interior mass distribution. But from the theory of equilibrium figures, it follows that only the homogeneous ellipsoids of *MacLaurin* exist in equilibrium. On the other hand, the surface of an equilibrium figure constructed of shells of equal density, and thus corresponding more to the real structure of the earth, is not an ellipsoid. Nevertheless, a layered structure of the interior mass of the level ellipsoid that approximates the actual situation, and sufficiently well reproduces the gravity field of the level ellipsoid, can be found (MORITZ 1968a). The maximum deviation between the level surfaces and the surfaces of equal density are on the order of f^2 only, and the differences in stress at the model remain considerably smaller than in the real earth. The level surface thus can also serve as a bounding surface for a *geophysical earth model* (MARUSSI et al. 1974).

There have been several attempts to construct earth models with a better fit to the geoid and the external gravity field than that provided by the level ellipsoid. A physical approximation consists of reference figures derived from truncated spherical harmonic expansions of the gravity potential of the earth: *level spheroids*. By assuming symmetry about the equator and truncating at the degree $l = 2$ (*Bruns'* spheroid) and $l = 4$ (*Helmert's* spheroid), we obtain surfaces of the fourteenth and twenty-second order respectively. The deviations from the rotational ellipsoid having the same axes are on the $O(f^2)$ for $l = 2$ and $O(f^3)$ for $l = 4$. From the harmonic coefficients $C_{2,2}$ and $S_{2,2}$ (3.102), a *triaxial ellipsoid* can also be calculated as a geometrical approximation to the geoid. This results in a difference of 70 m between the radii of the equatorial principal axes of inertia (corresponding to an equatorial flattening of 1/90 000), where the larger radius is directed to 345° longitude (BURŠA 1995).

These higher order earth models do not significantly reduce the deviations between the geoid and the level ellipsoid. On the other hand, computations related to these surfaces and their gravity fields become more complicated. Finally, they are not suitable as physical normal figures. Although, for instance, triaxial rotational ellipsoids exist as equilibrium figures (homogeneous ellipsoids of *Jacobi*), such an ellipsoid would yield a completely unnatural form when using the actual values for the angular velocity and mass of the earth.

4.2.2   The Normal Gravity Field of the Level Ellipsoid

The external gravity field of the level ellipsoid (normal gravity field) can be modeled by *closed formulas* in the system of *ellipsoidal coordinates* β,λ,u. The reduced latitude β and the geodetic longitude λ have been introduced in [4.1.1]. The third coordinate u is the semiminor axis of the ellipsoid with constant linear eccentricity ε, see (4.3), which passes through the point P (Fig. 4.7). From (4.20), and putting $\sqrt{u^2 + \varepsilon^2}$ for the semimajor axis, the transformation from the ellipsoidal coordinates to the Cartesian ones is given by

$$
\begin{pmatrix} \overline{X} \\ \overline{Y} \\ \overline{Z} \end{pmatrix} = u \begin{pmatrix} \sqrt{1+(\varepsilon/u)^2}\, \cos\beta\cos\lambda \\ \sqrt{1+(\varepsilon/u)^2}\, \cos\beta\sin\lambda \\ \sin\beta \end{pmatrix}.
\tag{4.35}
$$

For $\varepsilon = 0$, the β,λ,u-system with $\beta = 90° - \vartheta$ and $u = r$ degenerates into the system of spherical coordinates (2.14).
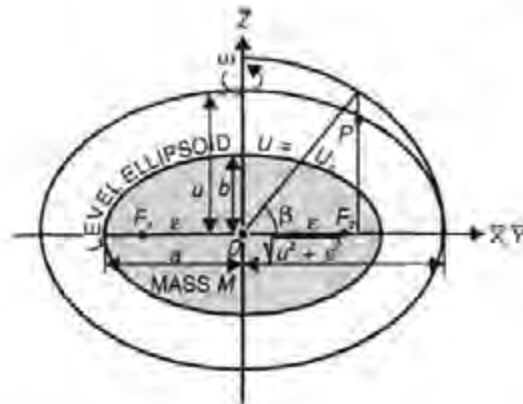
Fig. 4.7. Level ellipsoid and ellipsoidal coordinates

We denote the vector of *normal gravity* by γ and the normal gravity *potential* by U. In analogy to (3.43), we have

$$\gamma = \operatorname{grad} U . \tag{4.36a}$$

With respect to the global $\overline{X}, \overline{Y}, \overline{Z}$-system, γ is given in analogy to (3.72) by

$$\gamma = -\gamma \begin{pmatrix} \cos\varphi\cos\lambda \\ \cos\varphi\sin\lambda \\ \sin\varphi \end{pmatrix}. \tag{4.36b}$$

Corresponding to (3.42), U is composed of the gravitational potential V_E and the potential of the centrifugal acceleration Z_E:

$$U = V_E + Z_E . \tag{4.37}$$

The gravitational potential satisfies Laplace's differential equation (3.29) in the space exterior to the ellipsoid that contains the total mass.

By expressing Laplace's equation in ellipsoidal coordinates, we get a solution based on ellipsoidal harmonics. Adding the centrifugal potential (3.38), and taking both rotational symmetry and the condition of the ellipsoid surface as a level surface into account, we obtain a closed expression for the normal gravity potential (HEISKANEN and MORITZ 1967, p.64):

$$U = \frac{GM}{\varepsilon}\arctan\frac{\varepsilon}{u} + \frac{\omega^2}{2}a^2\frac{q}{q_0}\left(\sin^2\beta - \frac{1}{3}\right) + \frac{\omega^2}{2}\left(u^2 + \varepsilon^2\right)\cos^2\beta . \tag{4.38}$$

Here, q is an auxiliary quantity depending only on the geometric parameters ε and u. On the ellipsoid surface ($u = b$), it is denoted q_0:

$$q = \frac{1}{2}\left(\left(1 + 3\frac{u^2}{\varepsilon^2}\right)\arctan\frac{\varepsilon}{u} - 3\frac{u}{\varepsilon}\right), \quad q_0 = q_{u=b}. \tag{4.39}$$

Hence, in agreement with Stokes' theorem, cf. [4.2.1], the normal gravity potential is determined by four parameters (a, b, M, ω). It does not depend on the geodetic longitude. If one puts $u = b$ and $q = q_0$ in (4.38), then the potential of the *level ellipsoid* reads

$$U_0 = \frac{GM}{\varepsilon}\arctan\frac{\varepsilon}{b} + \frac{\omega^2}{3}a^2. \tag{4.40}$$

The normal gravity γ is perpendicular to the level ellipsoid, so that in accordance with (4.36), only the orthogonal component appears in the derivative of U (4.38). If the geodetic latitude φ is used instead of the reduced latitude β, then for the *normal gravity* on the *ellipsoid* we obtain the formula of *Somigliana* (1929):

$$\gamma_0 = \frac{a\gamma_a\cos^2\varphi + b\gamma_b\sin^2\varphi}{\sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}}. \tag{4.41a}$$

For numerical computations, the form

$$\gamma_0 = \gamma_a\frac{1 + k\sin^2\varphi}{\left(1 - e^2\sin^2\varphi\right)^{1/2}} \quad \text{with } k = \frac{b\gamma_b}{a\gamma_a} - 1 \tag{4.41b}$$

is more convenient (MORITZ 2000).

Here, the normal gravity, which depends on latitude only, is represented by the four parameters a, b, γ_a (normal gravity at the equator), and γ_b (normal gravity at the pole). The ellipsoidal parameters a, b, M, ω, γ, γ_b appearing in (4.38) and (4.41) are interrelated according to the theorem of *Pizetti*

$$2\frac{\gamma_a}{a} + \frac{\gamma_b}{b} = \frac{3GM}{a^2b} - 2\omega^2 \tag{4.42}$$

and the theorem of *Clairaut*

$$f + \beta = \frac{\omega^2 a}{\gamma_a}\left(1 + e'^2\right)^{-1/2}\left[1 + e'\frac{3\left(1 + \frac{1}{e'^2}\right)\left(1 - \frac{1}{e'}\arctan e'\right) - 1}{\left(1 + \frac{3}{e'^2}\right)\arctan e' - \frac{3}{e'}}\right]. \quad (4.43)$$

Thus, only four independent parameters are remaining. In (4.43), besides the second eccentricity e' and the geometric flattening f, there is also the *gravity flattening*

$$\beta = \frac{\gamma_b - \gamma_a}{\gamma_a}. \quad (4.44)$$

The abbreviation β is used for both the reduced latitude and the gravity flattening; confusion is not to be anticipated.

The normal gravity in the *exterior space* is obtained by partial differentiation of (4.38). Near the ellipsoid, a Taylor series expansion with respect to the ellipsoidal height is sufficient, see below.

Application of normal gravity field formulas, (4.38) to (4.43), is often facilitated by *series expansions* with respect to f, or some other quantity that describes the polar flattening.

We start with the spherical harmonic expansion of the gravitational potential. Due to the symmetry with respect to the rotational axis (tesseral terms = 0) and the equatorial plane (odd zonal terms = 0), we obtain, upon adding the centrifugal potential (3.96a) expressed in spherical coordinates, the *potential* of normal gravity in terms of Legendre polynomials, cf. [3.3.1],

$$U = \frac{GM}{r}\left(1 - \sum_{n=1}^{\infty}\left(\frac{a}{r}\right)^{2n} J_{2n} P_{2n}(\cos\vartheta)\right) + \frac{\omega^2}{2}r^2\sin^2\vartheta. \quad (4.45)$$

If P_2 is substituted from (3.83a), the expansion up to $n = 1$ (corresponding to $l = 2$) yields

$$U = \frac{GM}{r}\left(1 - \left(\frac{a}{r}\right)^2 J_2\left(\frac{3}{2}\cos^2\vartheta - \frac{1}{2}\right) + \frac{\omega^2}{2GM}r^3\sin^2\vartheta\right). \quad (4.46)$$

Solving for r and setting $U = U_0$ gives the *radius vector* to the level ellipsoid, where we have put $r = a$ on the right side:

$$r = \frac{GM}{U_0}\left(1 - J_2\left(\frac{3}{2}\cos^2\vartheta - \frac{1}{2}\right) + \frac{\omega^2 a^3}{2GM}\sin^2\vartheta\right). \qquad (4.47)$$

The *normal gravity* γ follows from the derivative of (4.46) with respect to r:

$$\gamma = \frac{GM}{r^2}\left(1 - 3\left(\frac{a}{r}\right)^2 J_2\left(\frac{3}{2}\cos^2\vartheta - \frac{1}{2}\right) - \frac{\omega^2}{GM}r^3\sin^2\vartheta\right). \qquad (4.48)$$

If we substitute either $\vartheta = 90°$ (equator) or $0°$ (pole) in (4.47) and (4.48), then we obtain either the semimajor axis a and the equatorial gravity or the semiminor axis b and the polar gravity of the ellipsoid. Using these values, the *geometric flattening* f (4.1a) and the gravity flattening β (4.44) may be computed according to

$$f = \frac{3}{2}J_2 + \frac{m}{2}, \quad \beta = -\frac{3}{2}J_2 + 2m. \qquad (4.49)$$

Here,

$$m = \frac{\omega^2 a^2 b}{GM} \approx \frac{\omega^2 a}{\gamma_a} \qquad (4.50)$$

is the ratio of the centrifugal acceleration to the normal gravity at the equator.

From (4.48) and (4.49), we arrive at an approximation to the theorem of *Pizetti* (4.42)

$$GM = a^2\gamma_a\left(1 - f + \frac{3}{2}m\right) \qquad (4.51)$$

and an approximation to *Clairaut's* theorem (4.43)

$$f + \beta = \frac{5}{2}m. \qquad (4.52)$$

Substituting (4.49) and (4.50) into (4.48), we obtain *Newton's* gravity formula, cf. [1.3.2] :

$$\gamma_0 = \gamma_a\left(1 + \beta\sin^2\varphi\right). \qquad (4.53)$$

If two γ_0 gravity values are known on the ellipsoid (gravity reduction problem) at different geographic latitudes φ, then γ_a and β may be computed from (4.53). With known values for the semimajor axis a and the angular velocity ω, (4.50) supplies the quantity m. Finally, Clairaut's theorem (4.52) yields the geometric flattening f, which thus can be determined from gravity values. Application of this principle to the real earth - that is, deriving geometric form parameters from physical quantities - leads to the gravimetric method of physical geodesy, cf. [6.5.1].

The relations above (linear in f, β, and m) may also be derived by series expansions of the closed formulas. They had already been found by *Clairaut* ("Theorie de la Figure de la Terre" 1743). The expansion up to terms of the order f^2 yields (IAG 1971)

$$f = \frac{3}{2}J_2 + \frac{m}{2} + \frac{9}{8}J_2^2 + \frac{15}{28}J_2 m + \frac{3}{56}m^2, \tag{4.54}$$

$$\beta = -f + \frac{5}{2}m - \frac{17}{14}fm + \frac{15}{4}m^2, \tag{4.55}$$

$$m = \frac{\omega^2 a^2 b}{GM}, \tag{4.56}$$

$$\gamma_0 = \gamma_a\left(1 + \beta\sin^2\varphi + \beta_1\sin^2 2\varphi\right), \quad \beta_1 = \frac{1}{8}f^2 - \frac{5}{8}fm. \tag{4.57}$$

One of the first applications of Clairaut's theorem was made by *Helmert* (1901). An adjustment to the gravity formula (4.57) of about 1400 free-air reduced gravity values yielded the parameters $\gamma_a = 9.7803\,\text{ms}^{-2}$ and $\beta = 0.005\ 302$, with a flattening of $f = 1/298.3$.

The harmonic coefficients of second and fourth degree may be computed from f and m as follows:

$$J_2 = \frac{2}{3}f - \frac{m}{3} - \frac{1}{3}f^2 + \frac{2}{21}fm, \quad J_4 = -\frac{4}{5}f^2 + \frac{4}{7}fm. \tag{4.58}$$

For today's accuracy requirements, an expansion up to $n = 3$ (corresponding to $l = 6$) is generally adequate. That is, the expansion has to include the terms of the order f^3 (COOK 1959). Developments up to the order f^5 have been given by CHEN (1982).

Near the earth's surface, a Taylor series expansion with respect to the ellipsoidal height h is sufficient for the derivation of the normal gravity in the *exterior space*:

$$\gamma = \gamma_0 + \left(\frac{\partial \gamma}{\partial h}\right)_0 h + \frac{1}{2}\left(\frac{\partial^2 \gamma}{\partial h^2}\right)_0 h^2 + \cdots . \tag{4.59}$$

The partial derivative $\partial \gamma / \partial h$ is obtained by applying Bruns' equation (3.71) to the exterior space:

$$\frac{\partial \gamma}{\partial h} = -2\gamma \bar{J} - 2\omega^2 , \tag{4.60}$$

where $\bar{J}$ is the mean curvature of the ellipsoid (4.19). A series expansion up to the order of f leads to the vertical component of the normal gravity gradient

$$\frac{\partial \gamma}{\partial h} = -2\frac{\gamma}{a}\left(1 + f + m - 2f \sin^2 \varphi\right). \tag{4.61}$$

The second derivative can be derived from a spherical approximation of γ, see (3.17). With

$$\gamma = \frac{GM}{r^2}, \quad \frac{\partial \gamma}{\partial r} = -2\frac{GM}{r^3} = -2\frac{\gamma}{r}$$

we obtain

$$\frac{\partial^2 \gamma}{\partial r^2} = \frac{6GM}{r^4} = 6\frac{\gamma}{r^2} . \tag{4.62}$$

Inserting the above into (4.59), with $r = a$ and $\gamma = \gamma_0$, leads to the normal gravity as a function of φ and h:

$$\gamma = \gamma_0\left(1 - \frac{2}{a}\left(1 + f + m - 2f \sin^2 \varphi\right)h + \frac{3}{a^2}h^2\right). \tag{4.63}$$

For higher altitudes, γ has to be derived by differentiating the normal gravity potential (4.38) either in the β,λ,u-system or, after corresponding transformation, in the φ,λ,h or the ϑ,λ,r-system, cf. HEISKANEN and MORITZ 1967, p.228.

With $\gamma = 9.81 \text{ms}^{-2}$ and $a = 6378$ km, we get $\partial\gamma/\partial h = -3.08\mu\text{ms}^{-2}/\text{m}$ and $\partial^2\gamma/\partial h^2 = 1.5 \times 10^{-6}\mu\text{ms}^{-2}/\text{m}^2$. More detailed numerical values are given in [4.3]. In gravity reductions the value $-3.086\mu\text{ms}^{-2}/\text{m}$ is used conventionally.

4.2.3   Geometry of the Normal Gravity Field

The geometry of the normal gravity field is represented by the spheropotential surfaces and the normal plumb lines.

The *spheropotential surfaces* are surfaces of constant normal gravity potential

$$U = U(\mathbf{r}) = \text{const.} \tag{4.64}$$

With the exception of the surface of the level ellipsoid ($U = U_0$), spheropotential surfaces deviate from ellipsoids and are not parallel to each other. The *normal plumb lines* intersect the spheropotential surfaces orthogonally. Due to the non-parallelism of the level surfaces, they are slightly curved in the plane of the meridian (Fig. 4.8).
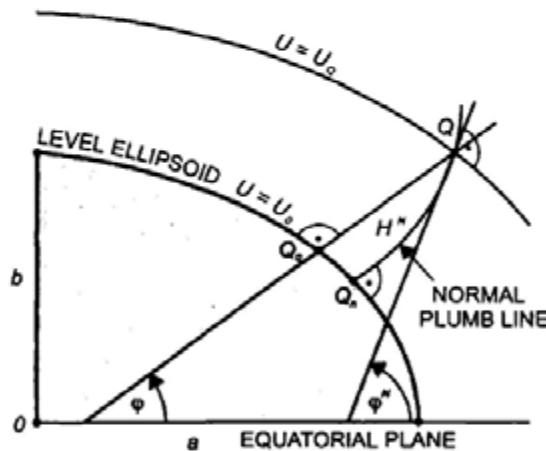
Fig. 4.8. Spheropotential surfaces, normal plumb line, normal height

In order to describe the geometry of the normal gravity field, "normal" geodetic coordinates φ^N, λ^N, U are introduced. They are defined in analogy to the "natural" coordinates Φ, Λ, W of the actual gravity field, cf. [3.2.3]. They refer to the point Q, which is related to the surface point $P(\Phi, \Lambda, W)$ by the conditions:

$$\varphi_Q^N = \Phi_P, \quad \lambda_Q^N = \Lambda_P, \quad U_Q = W_P. \tag{4.65}$$

The surface thus defined in a point-wise manner approximates the physical surface of the earth, with deviations less than 100 m and less than one arcmin respectively. This surface is called the *telluroid* (HIRVONEN 1960). It is not a level surface of the normal gravity field, but it resembles the earth's surface.

The *normal geodetic latitude* φ^N is the angle measured in the meridian plane between the equatorial plane of the ellipsoid and the direction of the normal plumb line. It differs from the geodetic latitude φ, introduced in [4.1.1], by the small angle $\delta\varphi^N$. This difference is a result of the plumb line curvature (Fig. 4.8), see below. *The normal geodetic longitude* λ^N is equal to the geodetic longitude λ. The *normal gravity potential* U relates the point Q to the level surface $U = U_Q$.

Instead of U, the potential difference $U_0 - U_Q$ to the level ellipsoid may be used. With the geopotential number $C = W_0 - W_P$ (3.104), and the condition $U_Q = W_P$, we also have $U_0 = W_0$. From this, we may derive the *normal height* H^N, already introduced in [3.4.3], by

$$H^N = \frac{U_0 - U_Q}{\overline{\gamma}} = \frac{W_0 - W_P}{\overline{\gamma}}. \tag{4.66}$$

Hence, H^N is defined as the distance between Q and the level ellipsoid measured along the normal plumb line. To a good approximation, H^N may be measured along the ellipsoidal normal passing through the surface point. According to (3.107), $\overline{\gamma}$ is the mean normal gravity between the ellipsoid and Q. Substituting γ from (4.63) into (3.107) and integrating yields

$$\overline{\gamma} = \gamma_0 \left(1 - \frac{1}{a} \left(1 + f + m - 2f \sin^2 \varphi \right) H^N \right). \tag{4.67}$$

Hence, $\overline{\gamma}$ may be computed rigorously in an iterative manner. Since C can be derived from measurements, the normal height can be determined without any hypothesis. Extending the normal heights downward from the earth's surface yields the *quasigeoid,* which is used as a reference surface for heights, cf. [3.4.3].

The normal geodetic coordinates $\varphi^N, \lambda^N, H^N$ have gained special importance for the direct determination of the physical surface of the earth according to the theory of Molodenski, cf. [6.5.1]. Normal heights have been introduced for a number of national height systems, cf. [3.4.3], [7.2].

The *curvature* of the *spheropotential surfaces* is described by the second derivatives of U, in analogy to the actual gravity field, see (3.57), (3.58). In the

local ellipsoidal system, the curvatures in the direction of the meridian and the parallel are given by

$$k_{\bar{x}}^N = -\frac{U_{\bar{x}\bar{x}}}{\gamma}, \quad k_{\bar{y}}^N = -\frac{U_{\bar{y}\bar{y}}}{\gamma}. \tag{4.68}$$

The geodesic *torsion* is zero due to the rotational symmetry of the level ellipsoid:

$$t_{\bar{x}}^N = -\frac{U_{\bar{x}\bar{y}}}{\gamma} = 0. \tag{4.69}$$

On the ellipsoid, the curvature is given by the principal radii of curvature M and N, see (4.13), (4.14):

$$k_{\bar{x}(0)}^N = \frac{1}{M}, \quad k_{\bar{y}(0)}^N = \frac{1}{N}. \tag{4.70}$$

Taking the rotational symmetry into account, and following (3.67), we get for the curvature of the projections of the *normal plumb line* onto the $\bar{x}, \bar{z}$ and the $\bar{y}, \bar{z}$ -plane respectively:

$$\kappa_{\bar{x}}^N = -\frac{U_{\bar{x}\bar{z}}}{g}, \quad \kappa_{\bar{y}}^N = -\frac{U_{\bar{y}\bar{z}}}{g} = 0. \tag{4.71}$$

On the level ellipsoid, we have with (4.20)

$$U_{\bar{x}\bar{z}(0)} = -\left(\frac{\partial\gamma}{\partial\bar{x}}\right)_0 = -\left(\frac{\partial\gamma}{M\partial\varphi}\right)_0. \tag{4.72}$$

Introducing the derivative $\partial\gamma/\partial\varphi$ from (4.53) and inserting into (4.71) yields with sufficient approximation

$$\kappa_{\bar{x}}^N = \frac{\beta}{M}\sin 2\varphi \tag{4.72a}$$

with β = gravity flattening (4.44). For the change of the normal gravity along the meridian, we thus get

$$\left(\frac{\partial\gamma}{\partial\bar{x}}\right)_0 = \gamma_0\frac{\beta}{M}\sin 2\varphi = 8.2\sin 2\varphi \text{ ns}^{-2}, \tag{4.72b}$$

which corresponds to 8.2 nms^{-2}/m at $\varphi = 45°$. Together with the relation $U_{\bar{z}\bar{z}} = -\partial\gamma/\partial\bar{z}$, (4.68) to (4.72) completely define the Marussi tensor (3.69) for the normal gravity field. According to (3.75), the differential transformation from the local to the global geodetic system is also provided by the curvature parameters.

Finally, we derive the differences between the geodetic coordinates φ, λ and the normal geodetic coordinates φ^N, λ^N:

$$\varphi = \varphi^N + \delta\varphi^N, \quad \lambda = \lambda^N. \tag{4.73}$$

From (4.71) and (4.72) we obtain

$$d\varphi^N = -\int_0^{H^N} \kappa_{\bar{x}}^N dH^N = -\frac{\beta}{M}\sin 2\varphi H^N.$$

And with $\beta = 0.0053$ and $M \approx a = 6371$ km we get

$$\delta\varphi^N = -0.000\,17''\sin 2\varphi H^N, \tag{4.74}$$

where H^N is in meters.

4.3   Geodetic Reference Systems

Geodetic reference systems provide numerical values for the parameters of a geodetic earth model. The systems are recommended by the International Union of Geodesy and Geophysics (IUGG) and represent the best parameter values for a designated epoch. The systems generally serve as a standard over a long time span for geodesy and related disciplines such as astronomy, cartography, geophysics, engineering, and navigation. Actual best values are determined at shorter time intervals and published by the International Association of Geodesy, cf. [6.8.1].

In the 19th and early 20th century, the geometric parameters of reference ellipsoids were derived from various terrestrial data sets and then introduced as a reference for national geodetic surveys, cf. [1.3.3], [7.1.2]. Normal gravity formulae referred to these ellipsoids have been derived since about 1900 and used for national gravimetric surveys. These regional or local reference systems

may be regarded as precursors of the present global systems, based on the theory of the level ellipsoid.

All reference systems are supposed to be geocentric, with the Z-axes coinciding with the earth's axis of rotation and the direction of the X-axis pointing to the Greenwich meridian. While the earlier reference systems may have large deviations from the geocentric system, recent reference systems agree at the "cm"-order. The orientation of geodetic systems with respect to the earth is described by the "Geodetic Datum", cf. [6.2.2].

Geodetic reference systems based on the theory of the level ellipsoid were first introduced in 1924/1930. At the IUGG General Assembly in Madrid 1924, *Hayford's* ellipsoid was introduced as the *International Ellipsoid*, with the parameters

$$a = 6378388 \,\mathrm{m}, \quad f = 1/297.0. \tag{4.75a}$$

The General Assembly in Stockholm (1930) adopted the gravity formula established by *G. Cassinis* for Hayford's ellipsoid:

$$\gamma_0 = 9.780\,49\left(1 + 0.005\,2884\sin^2\varphi - 0.000\,0059\sin^2 2\varphi\right)\mathrm{m\,s^{-2}}. \tag{4.75b}$$

This corresponds to the normal gravity formula (4.57), assuming a level ellipsoid.

The geometric parameters a and f were calculated by *J. F. Hayford* (1909) from astrogeodetic observations in the U.S.A. *W. A. Heiskanen* (1928) determined the equatorial gravity from an adjustment of isostatically reduced gravity values. The international reference system of 1924/1930 is thus defined by the four parameters a, f, γ_a, ω. The corresponding ellipsoid has been applied in numerous geodetic surveys; also, the normal gravity formula has found broad acceptance.

At the General Assembly of the IUGG in Luzern (1967), the 1924/1930 reference system was replaced by the *Geodetic Reference System 1967* (GRS67), see IAG (1971). It was defined by the following parameters:

$$a = 6\,378\,160\,\mathrm{m}, \quad GM = 398\,603\times10^9\,\mathrm{m^3\,s^{-2}}, \quad J_2 = 1082.7\times10^{-6}. \tag{4.76a}$$

The angular velocity of the earth's rotation

$$\omega = 7.292\,115\,146\,7\times10^{-5}\,\mathrm{rad\,s^{-1}}, \tag{4.76b}$$

not mentioned in the IUGG resolution, was adopted as the fourth parameter. The reference ellipsoid corresponding to this definition was declared a level ellipsoid.

The calculation of the semimajor axis was based on astrogeodetic observations collected over the entire earth, which were transformed into a uniform system by the inclusion of gravimetric data. Observations of space probes yielded the geocentric gravitational constant, which includes the mass of the atmosphere. The dynamic form factor was derived from the orbit perturbations of artificial satellites, and the angular velocity was adopted from astronomy. The GRS67 has been used especially for scientific problems and for a number of recently established geodetic networks.

At the IUGG General Assembly in Canberra (1979), the *Geodetic Reference System 1980* (GRS80) was introduced. It is also based on the theory of the geocentric equipotential ellipsoid, with the defining constants (MORITZ 2000):

$$
\left.
\begin{aligned}
&a = 6378137\,\text{m} && \text{equatorial radius of the earth} \\
&GM = 398\,600.5 \times 10^9\,\text{m}^3\,\text{s}^{-2} && \text{geocentric gravitational constant of the earth} \\
& && \text{(including the atmosphere)} \\
&J_2 = 1\,082.63 \times 10^{-6} && \text{dynamical form factor of the earth,} \\
& && \text{(excluding the permanent tidal deformation)} \\
&\omega = 7.292\,115 \times 10^{-5}\,\text{rad}\,\text{s}^{-1} && \text{angular velocity of the earth.}
\end{aligned}
\right\}
\quad (4.77\text{a})
$$

With $M_{atm} = 0.88 \times 10^{-6}\,\text{M}$, we have $GM_{atm} = 0.35 \times 10^9\,\text{m}^2\,\text{s}^{-2}$.

The system is consistent with the 1976 IAU system of astronomical constants, cf. [2.4.2].

With respect to the orientation, it is stated that the minor axis of the reference ellipsoid be parallel to the direction defined by the Conventional International Origin and that the primary meridian be parallel to the zero meridian of the BIH adopted longitudes, cf. [2.5.2].

The equatorial radius has been derived from laser distance measurements to satellites, satellite altimetry, and Doppler positioning, with an uncertainty of ±0.5 m. The calculation of the geocentric gravitational constant was based on space probes and lunar and satellite laser data $\left(\pm 0.1 \times 10^{-9} \right)$, while the value for the dynamic form factor was taken from global gravity models $\left(\pm 5 \times 10^{-9} \right)$.

Numerical values for derived parameters include (MORITZ 2000):

geometric constants, cf. [4.1.1], [4.1.2]:

$$
\left.\begin{array}{ll}
b = 6\ 356\ 752.3141\ \text{m} & \text{semiminor axis} \\
\varepsilon = 521\ 854.0097\ \text{m} & \text{linear eccentricity (4.3)} \\
c = 6\ 399\ 593.6259\ \text{m} & \text{polar radius of curvature (4.16)} \\
e^2 = 0.006\ 694\ 380\ 022\ 90 & \text{first eccentricity (e) (4.1b)} \\
e'^2 = 0.006\ 739\ 496\ 775\ 48 & \text{second eccentricity (e´) (4.1c)} \\
f = 0.003\ 352\ 810\ 681\ 18 & \text{flattening (4.1a)} \\
1/f = 298.257\ 222\ 101 & \text{reciprocal flattening} \\
G = 10\ 001\ 965.7293\ \text{m} & \text{meridian quadrant (4.21a)}
\end{array}\right\} \quad (4.77b)
$$

physical constants, cf. [4.2.2]:

$$
\left.\begin{array}{ll}
U_0 = 62\ 636\ 860.850\ \text{m}^2\ \text{s}^{-2} & \text{normal potential at ellipsoid} \quad (4.40) \\
J_4 = -0.000\ 002\ 370\ 912\ 22 & \text{spherical harmonic coefficient (4.45)} \\
J_6 = 0.000\ 000\ 006\ 083\ 47 & \text{spherical harmonic coefficient (4.45)} \\
J_8 = -0.000\ 000\ 000\ 014\ 27 & \text{spherical harmonic coefficient (4.45)} \\
m = 0.003\ 449\ 786\ 003\ 08 & (4.50) \\
\gamma_a = 9.780\ 326\ 7715\ \text{m s}^{-2} & \text{normal gravity at equator (4.41)} \\
\gamma_b = 9.832\ 186\ 3685\ \text{m s}^{-2} & \text{normal gravity at pole} \quad (4.41) \\
\beta = 0.005\ 302\ 440\ 112 & (4.44) \\
k = 0.001\ 931\ 851\ 353 & (4.41b)
\end{array}\right\} \quad (4.77c)
$$

Normal gravity can be computed by the closed formula (4.41) or the series expansion

$$
\gamma_0 = \gamma_a \left(\begin{array}{l} 1 + 0.005\ 279\ 0414\sin^2\varphi + 0.000\ 023\ 2718\sin^4\varphi \\ + 0.000\ 000\ 1262\sin^6\varphi + 0.000\ 000\ 0007\sin^8\varphi \end{array}\right), \quad (4.78a)
$$

which is accurate to $10^{-3}\,\mu\text{m s}^{-2}$. The conventional series (4.57) has an accuracy of only $1\mu\text{m}^{-2}$:

$$
\gamma_0 = 9.780\ 327\left(1 + 0.005\ 3024\sin^2\varphi - 0.000\ 0058\sin^2 2\varphi\right)\text{m s}^{-2}. \quad (4.78b)
$$

Inserting the values for the GRS80 into (4.63) yields the change of normal gravity with height:

$$
\gamma = \gamma_0 - \left(3.0877\times10^{-3} - 4.3\times10^{-6}\sin^2\varphi\right)h + 0.72\times10^{-6}h^2\ \text{m s}^{-2} \quad (4.79)
$$

with h in km. A development accurate to $10\,\text{nm s}^{-2}$ for heights up to 10 km is given by WENZEL (1989).

According to the definition of GM, γ_0 refers to the total mass of the earth including the atmosphere. If normal gravity values are required on the ellipsoid, or within the range of the atmosphere, the effect of the air masses above the calculation point must be subtracted from γ_0. The corresponding reduction

amounts to $-8.7\,\mu\mathrm{m\,s^{-2}}$ $(h=0)$, $-4.7\,\mu\mathrm{m\,s^{-2}}$ $(h=5\,\mathrm{km})$, and $-0.1\,\mu\mathrm{m\,s^{-2}}$ $(h=30\,\mathrm{km})$, ECKER and MITTERMAYER (1969).

5 Methods of Measurement

Modeling of geodetic parameters is based on observations taken on the earth's surface and in it's exterior space. Different measurement methods are available, delivering geometric or physical quantities. Geometric methods rely primarily on electromagnetic waves and thus are influenced by atmospheric refraction [5.1]. The measurement methods may be divided into

- observations employing artificial satellites as targets (including the moon) or sensors: satellite observations [5.2],
- observations to fixed stars and extragalactic radio sources: geodetic astronomy [5.3],
- terrestrial gravity and gravity gradient measurements: gravimetry [5.4],
- determination of coordinate differences between points on the surface of the earth: terrestrial geodetic measurements [5.5].

The measurement methods depend on available technology, where electronics governs data collection and online data-processing (KAHMEN 1978, SCHLEMMER 1996). Satellite techniques now dominate global and regional surveys, while terrestrial methods are used for interpolation in space and time. Accuracy limits are governed by calibration errors and elimination or reduction of disturbing influences such as temperature, atmospheric pressure, magnetic field variations, microseismicity, atmospheric refraction, and local instabilities. Thus the inherent precision of the respective technique may deteriorate by a factor of two to three, or more.

5.1 Atmospheric Refraction

In practically all geodetic measurements, electromagnetic waves serve as signal carriers; this includes the methods of satellite and terrestrial geodesy as well as geodetic astronomy. From the broad spectrum of electromagnetic waves, the visible light (380 to 780 nm, corresponding to 7.9 to $3.8 \cdot 10^{14}$ Hz), the near infrared (up to 1 μm), and the microwave parts (1mm to 1 m, corresponding to 300 GHz to 300 MHz) are used. When traveling through the atmosphere, the waves experience changes in velocity and curvature of the path, depending on the physical state of the atmosphere [5.1.1]. Signal propagation is different in the troposphere and the ionosphere and has to be treated separately [5.1.2], [5.1.3].

Different methods have been developed in order to eliminate or reduce the effects of atmospheric refraction on geodetic measurements. These include

instrument design, observation methodology, and the use of atmospheric models based on data collected on the earth's surface and in space. The individual strategies will be discussed in the chapters that pertain to measurement methods, see also BRUNNER (1984a), DE MUNCK and SPOELSTRA (1992).

Acoustic waves are employed for positioning on, and mapping of, the ocean floor and the bottom of rivers and lakes. The propagation of these waves through water differs from their behavior in vacuum and depends on water properties, cf. [5.5.2].

5.1.1 Fundamentals

According to *Fermat's principle*, the path s of an electromagnetic wave is determined by the condition of a minimum travel time Δt of the wave (MORITZ and HOFMANN-WELLENHOF 1993, p.158):

$$\Delta t = \int_{path} dt = \int_{path} \frac{ds}{v} = \min .\tag{5.1}$$

The velocity v differs from the velocity in vacuum c (2.2) by the *index of refraction n* (also called refractive index):

$$n = \frac{c}{v} .\tag{5.2}$$

For a gaseous medium, $n > 1$ is proportional to the density of the gas. If the medium is dispersive for a certain spectral domain, n also depends on the wavelength: *dispersion*. An average value for n near the earth's surface is 1.0003. Instead of n, the *refractivity*

$$N = (n-1) \cdot 10^6\tag{5.3}$$

is frequently used.

Inserting (5.2) into (5.1) yields

$$\Delta t = \frac{1}{c} \int_{path} n \, ds = \min .\tag{5.4}$$

By setting

$$n \, ds = d\overline{s} ,$$

(5.4) can also be expressed as a minimum condition for the "electromagnetic" path length (Fig. 5.1):

$$\bar{s} = \int_{path} d\bar{s} = \int_{path} n \, ds = \min .$$ (5.5)

Solving the variational problem (5.5) yields the path $\bar{s}$, but requires the knowledge of n along the path.
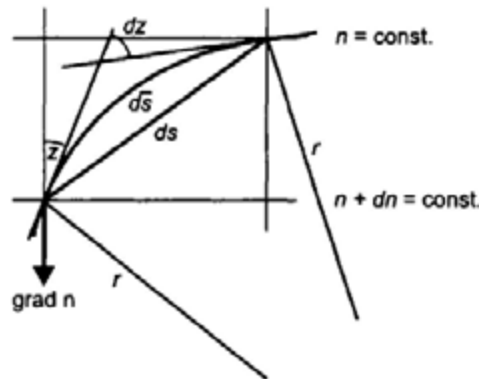
Fig. 5.1. Ray bending in the atmosphere

The effect of refraction on *distance* is given by the difference between the actual path length $\bar{s}$ and the straight-line s (chord):

$$\bar{s} - s = \int_{path} n \, ds - \int_0^s ds = \int_0^s (n-1) ds + \left(\int_{path} n \, ds - \int_0^s n \, ds \right).$$ (5.6)

The first term on the right side accounts for the difference in length due to the longer travel time in the atmosphere, while the second term represents the effect of the bending of the ray (JANES et al. 1991).

The refraction effect on *curvature* can be estimated by assuming that the air density is stratified horizontally (Fig. 5.1). *Snell's law* describes the bending of the ray as it passes through layers of varying refractive index, which corresponds to Fermat's principle:

$$n \sin z = \text{const.}$$ (5.7a)

Or for two points P_1 and P_2:

$$n_1 \sin z_1 = n_2 \sin z_2 . \tag{5.7b}$$

Under the above assumption, the angle between the normal to the surface $n = $ const. and the tangent to the ray with curvature $1/r$ is the zenith angle z. Differentiation of (5.7) yields

$$\sin z \, dn + n \cos z \, dz = 0 .$$

With

$$dn = (\mathrm{grad}\ n) \cdot d\mathbf{s} = |\mathrm{grad}\ n| \cos z \, ds ,$$

we obtain the curvature

$$\frac{1}{r} = \frac{dz}{ds} = -\frac{|\mathrm{grad}\ n|}{n} \sin z . \tag{5.8}$$

By separating the horizontal and the vertical component of grad n, we get the curvatures of the ray projected into the horizontal and the vertical planes. The corresponding effects on horizontal and vertical angles are called *horizontal* (lateral) and *vertical* refraction, respectively.

Horizontal refraction is about one to two orders of magnitude less than vertical refraction. Neglecting it yields a simplified formula for the curvature of *vertical refraction*:

$$\frac{1}{r} = -\frac{1}{n}\frac{dn}{dh} \sin z , \tag{5.9a}$$

where h is the height, cf. [4.1.3].

In *terrestrial* geodetic measurements, we have $n \approx 1$ and $z \approx 90°$, which leads to

$$\frac{1}{r} = -\frac{dn}{dh} . \tag{5.9b}$$

Instead of $1/r$, the *coefficient of refraction k* is often used. It is defined as the ratio between the radius of the earth R and the curvature radius r:

$$k = \frac{R}{r} = -R\frac{dn}{dh} . \tag{5.10}$$

The vertical *refraction angle* δ is the effect of refraction on observed zenith angles (Fig. 5.2). It results from integrating $1/r$ resp. dn/dh along the path:

$$\delta = \frac{1}{s}\int_0^s (s - s_i)\frac{dn}{dh}ds .\qquad (5.11a)$$
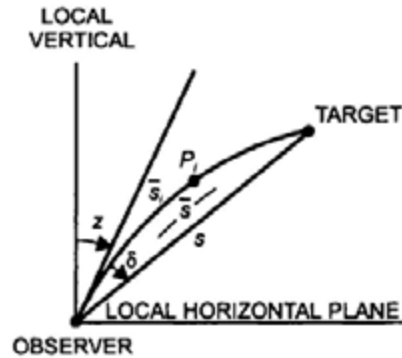
Fig 5.2. Vertical refraction

Here the local vertical gradient of n is weighted according to the distance from the observer; values from closer distances receive a greater weight. For a spherical arc ($r = $ const.), and taking (5.10) into account, (5.11a) reduces to

$$\delta = \frac{k}{2R}s .\qquad (5.11b)$$

In most geodetic applications, the signal is transferred by a modulation of the carrier wave. This can be regarded as a superposition of a group of waves with different frequencies. While *phase velocity* v_{ph}, introduced in (5.2), refers to the monochromatic carrier wave, the center of a short wave group (signal energy) travels with the *group velocity*

$$v_{gr} = v_{ph} - \lambda\frac{dv_{ph}}{d\lambda} .\qquad (5.12)$$

In a dispersive medium, we have $n = n(\lambda)$ and $dv_{ph}/d\lambda \neq 0$ (LEICK 1995). Taking (5.2) into account delivers the corresponding *group refractive index*

$$n_{gr} = n_{ph} - \lambda\frac{dn_{ph}}{d\lambda} = n_{ph} + f\frac{dn_{ph}}{df} ,\qquad (5.13)$$

with f = frequency, cf. [5.1.2], [5.1.3].

For *standard air* (temperature 273.15 K, air pressure 1013.25 hPa, humidity 0.0 hPa, CO_2 content 0.0375 %), the *phase refractivity* may be calculated as follows (IAG resolution, Gen. Ass. Birmingham 1999):

$$N_{ph} = (n_{ph} - 1)10^6 = 287.6155 + \frac{1.62887}{\lambda^2} + \frac{0.01360}{\lambda^4}, \qquad (5.14)$$

where λ is the carrier wave length in μm, and n_{ph} is the corresponding phase refractive index. The *group refractivity* is given by

$$N_{gr} = (n_{gr} - 1)10^6 = 287.6155 + \frac{4.88660}{\lambda^2} + \frac{0.06800}{\lambda^4}, \qquad (5.15)$$

with n_{gr} = group refractive index.

According to (5.6) and (5.11), the refraction effect on distances and angles depends on the index of refraction and its gradient along the path of the ray, which behave differently in the troposphere than in the ionosphere.

5.1.2 Tropospheric Refraction

The *troposphere* is the lower layer of the atmosphere. It extends to a height of about 9 km at the poles and 16 km at the equator. Weather processes take place in tropospheric region, where nearly 90% of the atmospheric masses are concentrated. The *tropopause* as a boundary layer separates the troposphere from the *stratosphere,* which extends to about 50 km. The troposphere, tropopause, and stratosphere are electronically neutral. The index of refraction n depends on temperature T, pressure p and humidity e. For visible light, the troposphere behaves as a dispersive medium, cf. [5.1.1]. The refractive index decreases with height and becomes nearly 1 at about 40 km. Tropospheric refraction is the combined effect from the ground to this "effective" height. Above 50 km the atmosphere is ionized, cf. [5.1.3].

The *meteorological parameters T, p, e* not only depend strongly on height but also on latitude, land/ocean distribution, topography, vegetation, and local conditions. These variables produce large to small-scale anomalies of n. Variations with time are of long-term, seasonal, daily, and turbulent character. Rapid fluctuations are especially pronounced close to the earth's surface, up to 10 to 30 m above the ground.

Temperature T decreases in the troposphere almost linearly with height h according to

$$dT/dh \approx -0.0065°\text{C/m}, \tag{5.16a}$$

followed by a slight increase in the stratosphere. Horizontal temperature gradients may reach a few °C/100 km. Close to the ground, temperature variations are pronounced, including temperature inversion during night time and convection at noon.

Air pressure p decreases exponentially with height. Assuming hydrostatic equilibrium, the vertical pressure gradient depends on density ρ and gravity g. Near the *surface* of the earth, this leads to

$$dp/dh = -\rho g = -0.034\frac{p}{T} = -0.12\,\text{h Pa/m} \tag{5.16b}$$

at standard conditions ($T = 288$ K, $p = 1013$ hPa).

Humidity is rather irregularly distributed and concentrated in a layer of a few km above ground, where strong variations also occur with time. Humidity is measured by the water vapour pressure e, which is about 10 to 20 hPa at mid-latitudes close to the surface. It tends to decrease with height, with a mean value of

$$de/dh = -0.0035\,\text{h Pa/m} \tag{5.16c}$$

near the ground.

Empirical relations have been derived between the index of refraction and the meteorological parameters for both light and microwaves (BOMFORD 1980, p.42, LANGLEY 1998).

With atmospheric conditions different from the standard air, cf. [5.1.1], the *group refractivity* of visible *light* and near *infrared* waves in ambient moist air is (IAG resolution, Gen. Ass.Birmingham 1999):

$$N_l = (n_l - 1)10^6 = \frac{273.15}{1013.25}\frac{p}{T}N_{gr} - 11.27\frac{e}{T}, \tag{5.17}$$

with T in Kelvin, p and e in hPa. Equation (5.17) is also valid for unmodulated light with the corresponding phase refractivity (5.14).

The refractivity of *microwaves* (independent of the wavelength) is given by the formula of Thayer (1974)

$$N_m = (n_m - 1)10^6 = 77.60\frac{p}{T} - 13\frac{e}{T} + 3.78 \cdot 10^5\frac{e}{T^2}, \tag{5.18}$$

which is practically identical with the formula of Essen and Froome (IAG resolution, Gen. Ass. Berkeley 1963).

The first term on the right side of (5.18) represents a "dry" component of the refractivity. It contributes about 90% to the total tropospheric refraction in the lower 15 km and can be modeled from surface pressure, assuming hydrostatic equilibrium. The "wet" component, as expressed by the terms depending on e (especially the last one), is extremely difficult to model but approaches zero at around a height of 10 km.

In order to keep the error in the index of refraction less than 10^{-6}, the meteorological parameters in (5.17) and (5.18) have to be determined to about ±1°C for temperature, ±3.5 hPa for pressure, and ±25 hPa (light) resp. ±0.2 hPa (microwaves) for humidity.

Differentiating (5.17) and (5.18) with respect to the height h yields the dependence of the *curvature* on the meteorological parameters. Neglecting minor terms and taking (5.16b) into account, we obtain for the surface near layers

$$\frac{dN_l}{dh} = -78\frac{p}{T^2}\left(0.034 + \frac{dT}{dh}\right) - \frac{11}{T}\frac{de}{dh} \tag{5.19a}$$

for *light*. For *microwaves*, the last term on the right side (wet component) changes to

$$+\frac{3.7\cdot 10^5}{T^2}\frac{de}{dh}. \tag{5.19b}$$

Thus according to (5.10), the *coefficient of refraction k* has to be determined for light with an accuracy of ±2°C in temperature, ±6 hPa in air pressure, and ±0.0002°C/m in the temperature gradient in order to achieve a relative accuracy of 1%. For microwaves, the admissible errors may be two times larger. The gradient of the water vapor pressure should be determined with ±0.005 hPa/m for light and ±0.0001 hPa/m for microwaves. Hence, the most critical parameters are the vertical gradients of temperature and, especially for microwaves, of the water vapor pressure. According to (5.11b), an error of 1% in k would produce an error in the refraction angle of ±0.2" over a distance of 10 km and ±0.4" over 25 km.

In the layers close to the ground, the strong variations of the meteorological parameters in space and time lead to corresponding changes in the *coefficient of refraction*, with pronounced seasonal and day/night variations (HOEPCKE 1966). Under average daytime conditions with a clear sky, and for heights between 40 m and 100 m above the ground, we have for *light*

$$k_l = 0.13 \text{ or } r_l = 8R, \tag{5.20a}$$

and for *microwaves*

$$k_m = 0.25 \text{ or } r_m = 4R.$$ (5.20b)

Tropospheric models generally assume concentric spherical layers and azimuthal symmetry and neglect variations with time. Global models are provided by standard atmospheres in the form of vertical profiles for temperature, pressure, and density. The U.S. standard atmosphere (1976) approximates mean mid-latitude conditions for dry air. Latitudinal and seasonal departures are given by supplements, NOAA (1966, 1976).

More refined refraction models have been developed for the reduction of geodetic measurements; these models employ observed surface data as input (HOPFIELD 1969, SAASTAMOINEN 1972/1973, NEILL 1996). A "mapping function" takes into account the increase in refraction with larger zenith angles z. A simple approximation is given by $1/\cos z$ for elevation angles that are not too small.

The path delays that GPS (Global Positioning System) *signals* experience when passing through the atmosphere, cf. [5.2.5], can be exploited for tropospheric and ionospheric monitoring (DAVIS et al. 1996). By separating the "dry" component from the *tropospheric* signal delay, the integrated perceptible *water vapor* can be estimated from the "wet" component (Bevis et al. 1992). Permanent GPS networks, cf. [7.3.1], supply this information on global and regional scales, with high temporal resolution.

Atmospheric sounding by *radio occultation* is possible with spaceborne GPS receivers installed on Low Earth Orbiters (LEO's), YUNK and MELBOURNE (1996). Here, the GPS signal is tracked after rising or before setting of the GPS satellite. The observed Doppler shift induced by atmospheric bending is used for constructing profiles of atmospheric density, pressure, temperature, and water vapor content, and evaluated for developing tropospheric models.

5.1.3 Ionospheric Refraction

As part of the higher atmosphere, the ionosphere is characterized by the presence of free, negatively-charged electrons and positive ions. Ionization is caused primarily by the impact of solar ultraviolet radiation and consequently depends on the density of the atmospheric gas and the intensity of the radiation. The ionosphere covers the region between 50 km and 1500 km above the earth, with a maximum electron density at a height of 200 to 300 km.

The ionosphere acts like a mirror at frequencies below 30 MHz. Radio waves of higher frequencies pass through the ionosphere but experience frequency-dependent effects (dispersive medium). Measurements to targets above the ionosphere are also affected by the electron concentration in the plasmasphere,

which extends up to a height of several earth radii above the equator and does not exist at the poles (KLOBUCHAR 1991, WANNINGER 1995).

The *index of refraction* depends primarily on the number N_e of electrons per m^3: *electron density*. As a first order approximation, the *phase* refractive index is given by

$$n_{ph} = 1 - K \frac{N_e}{f^2},$$ (5.21)

with the constant $K = 40.28$ m^{3}s^{-2} and f = frequency. Higher terms of the order $1/f^3$ and $1/f^4$ also depend on the intensity of the earth's magnetic field and the direction of the signal propagation. In daytime, N_e (el/m^3) varies between about $10^8 \dots 10^{10}$ (60 to 90 km) over 10^{11} (105 to 160 km) to $10^{11} \dots 10^{12}$ (160 to 180 km) and 10^{12} (300 to 400 km).

As seen from (5.2) and (5.21), the phase velocity is larger than the velocity of light in vacuum, which corresponds to a larger wavelength of the signal compared to vacuum. Since signal propagation follows the group velocity, we insert (5.21) into (5.13) and obtain the group refractive index

$$n_{gr} = 1 + K \frac{N_e}{f^2}.$$ (5.22)

Inserting (5.21) resp. (5.22) into (5.6) delivers the difference between the electromagnetic path length $\overline{s}$ and the straight-line connection s. This yields for carrier phase (n_{ph}) and for range (n_{gr}) observations respectively:

$$(\overline{s} - s)_{ph} = -(\overline{s} - s)_{gr} = -\frac{K}{f^2} \int_0^s N_e \, ds,$$ (5.23)

where the small effect of the path's bending has been neglected.

The integral of the electron density along the path is called *total electron content* (TEC):

$$TEC = \int_0^s N_e (s) \, ds.$$ (5.24)

It gives the number of electrons along the signal path in a cylindrical column with a cross section of 1 m^2. Its unit is 1 TECU $= 10^{16}$ electrons/m^2. For a spherically-layered ionosphere, we may introduce the electron content along a

vertical column of height h and relate it to the TEC along the path by an oblique factor F ("mapping function"):

$$\text{TEC} = F \int_0^h N_e(h)\, dh. \tag{5.25}$$

For $z < 70°$, we have $F \approx 1/\cos z_i$, with z_i = zenith angle at the subionospheric point P_i (Fig. 5.3). P_i is located at the "mean height" h_i of the ionosphere (single-layer model), with, e.g., $h_i = 350$ km. z_i can be calculated from h_i and the zenith angle z measured from the ground:

$$\sin z_i = \frac{R}{R + h_i} \sin z, \tag{5.26}$$
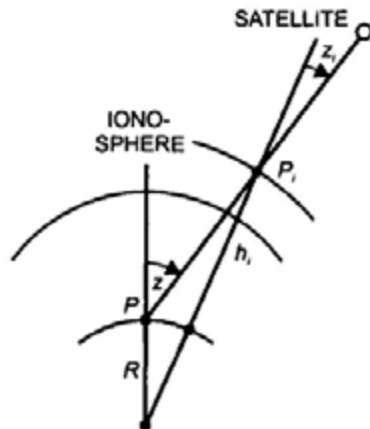
R = radius of the earth.

Fig. 5.3. Ionospheric refraction

The electron content in the ionosphere varies strongly with time. There are pronounced variations of daily, seasonal, and about an 11-year (solar activity cycle) period. Superimposed on to these more regular variations are irregular disturbances. Short-term scintillations occur primarily in the equatorial zones but also in the polar and auroral regions (magnetic storms). Traveling ionospheric disturbances of wavelike structure proceed with horizontal speeds between 100 and 1000 m/s at scales of some 10 to 1000 km and at periods from several minutes to a few hours.

Ionospheric models describe the distribution of N_e in space and time; they are based on the dependency of the ionospheric state on the position of the sun. Assuming a spherical shell distribution, they provide a smoothly varying TEC along vertical profiles and a mapping function for inclined signal propagation.

Among these models are the International Reference Ionosphere (IRI) 1995 and the GPS-broadcast model (KLOBUCHAR 1991). These models may deviate from reality by some 10%, due to the ionospheric disturbances. Refined models include observed ionospheric parameters and sunspot numbers (BROWN et al. 1991, KLEUSBERG 1998). The dispersion effect, on the other hand, allows us to eliminate most of the ionospheric refraction effects by employing two frequencies for signal propagation, cf. [5.2.4], [5.2.5], [5.2.7].

The ionospheric delay of GPS signals provides information on the structure and temporal behavior of the ionosphere. By analyzing the two carrier waves used to eliminate ionospheric refraction, cf. [5.5.2], the total electron content (TEC) along the line of sight from the receiver to the GPS satellite can be measured. The results of global GPS ground networks are used to generate global maps of TEC. Space-based GPS data, cf. [5.1.2], will significantly improve this ionospheric imaging, especially the vertical resolution (YUNK and MELBOURNE 1996).

5.2   Satellite Observations

Satellite geodesy utilizes artificial satellites and the moon as extraterrestrial targets or sensors. For a point mass earth model, the orbital motion of a satellite is described by Kepler's laws [5.2.1]. The actual gravitational field and non-gravitational forces create orbital "perturbations" [5.2.2]. Satellites used for geodetic applications differ in design, equipment, and orbit according to the respective observation technique and the mission purpose [5.2.3]. Classical measurement methods, introduced and employed from the 1960's to the 1980's, demonstrated the efficiency of satellite observations for establishing large-region geodetic control networks and gravitational field determination [5.2.4]. Today, the Global Positioning System (GPS) governs three-dimensional positioning at all scales [5.2.5], while laser distance-measurements primarily contribute to global reference networks [5.2.6]. By monitoring the ocean surface, satellite altimetry contributes to gravity field modeling [5.2.7], and high-resolution global gravity field recovery is expected from future satellite-to-satellite tracking and gravity gradiometry missions [5.2.8].

The theory of satellite orbits and satellite measurement methods are treated in textbooks and monographs on satellite geodesy, e.g., KAULA (1966), SCHNEIDER (1988), SEEBER (1993).

5.2.1 Undisturbed Satellite Motion

After the satellite has separated from the carrier, it begins its unrestrained revolution about the earth. We assume the gravitational point mass model, cf. [3.1.2], and neglect the mass of the satellite with respect to the earth's mass. If we also neglect perturbations of non-gravitational type and the effect of other

celestial bodies (two-body problem), Newton's second law of motion provides the *equation of motion* in the gravitational field:

$$\ddot{\mathbf{r}} = \text{grad } V = -\frac{GM}{r^2}\frac{\mathbf{r}}{r}. \tag{5.27}$$

r is the geocentric position vector of the satellite and $\ddot{\mathbf{r}} = d^2\mathbf{r}/dt^2$ its acceleration, M and V are the mass and the gravitational potential of the earth respectively, cf. (3.16). The integration of the vectorial second-order differential equation (5.27) introduces six integration constants, e.g., position and velocity at a given epoch.

The fundamental theory on the two-body problem is given by celestial mechanics (BROUWER and CLEMENCE 1961, KOVALEVSKY 1989, SCHNEIDER 1992/1993). Such works also address the fundamentals of orbit perturbations, orbit computation, and the treatment of three and multi-body problems.

Johannes Kepler (1571-1630) derived three laws of planetary motion from the astronomic observations collected by *Tycho de Brahe* (1546-1601). *Keplerian motion* provides a geometric description of the equation of motion for an undisturbed central motion.

According to Kepler's laws, the satellite moves in an elliptical orbit. One focal point of the ellipse, with semimajor axis a and first eccentricity e (not to be confused with the corresponding parameters of the earth ellipsoid), coincides with the center of mass of the earth. In the orbital system (Fig. 5.4), the position of the satellite is described by the distance r from the center of mass and the true anomaly v. The true anomaly is the geocentric angle between the directions to the satellite and perigee. Instead of v, the eccentric anomaly E can be used, with the relations

$$r = a\left(1 - e\cos E\right), \quad \tan v = \frac{\sqrt{1-e^2}\,\sin E}{\cos E - e}. \tag{5.28}$$

With Kepler's third law, the mean (angular) velocity

$$\bar{n} = \sqrt{\frac{GM}{a^3}} \tag{5.29}$$

is introduced, describing a mean orbital motion. The mean anomaly

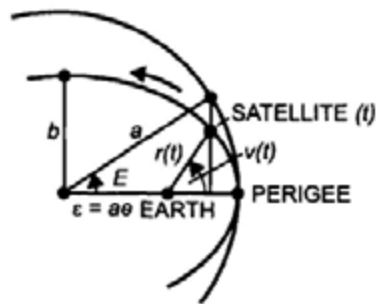$$\bar{M} = \bar{n}\left(t - T\right) \tag{5.30}$$

Fig. 5.4. Satellite orbital system

represents yet another parameter for describing the satellite's position in the orbit. It is generally preferred because it increases *linearly* with time. T is the epoch of the passage through the perigee, the closest approach to the earth. From $\bar{M}$, E can be computed iteratively using Kepler's equation:

$$\bar{M} = E - e \sin E .\tag{5.31}$$

The orbital system is transformed into the space-fixed equatorial system, cf. [2.4.1], by three rotations (Fig. 5.5). The right ascension of the ascending node Ω and the inclination i provide the orientation of the orbital plane in space. The argument of perigee ω orients the ellipse in the orbital plane. From the result of this transformation, we obtain the geocentric position vector (2.11) as a function

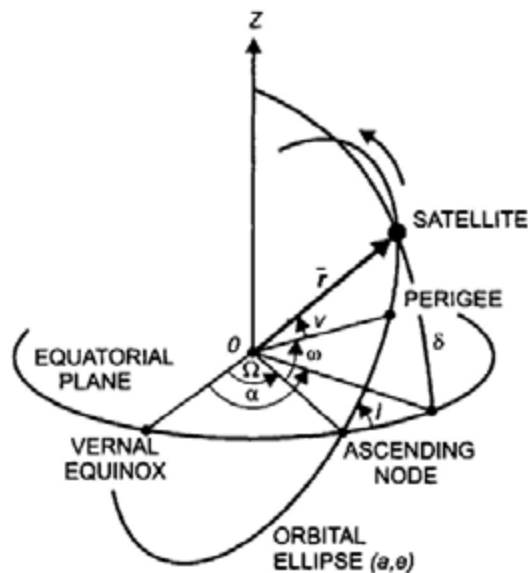*Fig. 5.5.* Orbital and equatorial system

of the six Keplerian elements a, e, Ω, i, ω, and v (or equivalently E, $\bar{M}$, or T):

$$\mathbf{r} = r \begin{pmatrix} \cos\delta\cos\alpha \\ \cos\delta\sin\alpha \\ \sin\delta \end{pmatrix} = r \begin{pmatrix} \cos(\omega+v)\cos\Omega - \sin(\omega+v)\sin\Omega\cos i \\ \cos(\omega+v)\sin\Omega - \sin(\omega+v)\cos\Omega\cos i \\ \sin(\omega+v)\sin i \end{pmatrix}, \quad (5.32a)$$

with

$$r = \frac{a(1-e^2)}{1+e\cos v}. \quad (5.32b)$$

The six Keplerian parameters completely describe the orbital motion of the undisturbed satellite. They correspond to the integration constants of the equation of motion (5.27) and are used for the approximation of satellite orbits.

5.2.2 Perturbed Satellite Motion

The actual orbit of a satellite departs from the Keplerian orbit due to the effects of various "disturbing" forces. This includes the non-spherical parts of the earth's gravitation, the gravitational effects of moon and sun, atmospheric air drag, and solar radiation pressure, among others. These disturbing forces cause variations in time in the orbital elements (orbital perturbations) of secular, long, and short-periodic nature. The actual orbit can be viewed as the envelope of Keplerian ellipses, which are given at each instant by the actual orbital elements (osculating ellipses).

In order to account for the complete *gravitation* of the earth, the gravitational potential of a spherically symmetric earth must be amended by the *perturbing potential R* (not to be confused with the *disturbing* potential [6.1.1]):

$$V = \frac{GM}{r} + R. \quad (5.33)$$

According to equations (3.89) to (3.91), R can be expressed by a spherical harmonic expansion of the gravitational potential V through the harmonic coefficients J_{lm}, K_{lm} $(l \geq 2)$. By substituting (5.33) into (5.27), the *equation of motion* now reads

$$\ddot{\mathbf{r}} = -\frac{GM}{r^2}\frac{\mathbf{r}}{r} + \text{grad } R. \quad (5.34a)$$

The spherical coordinates r, ϑ, λ (2.14) used in the expansion of V can be replaced by the orbital elements according to (5.28) and (5.32), see also Fig. 5.4. R then is described by the time-variable Keplerian orbital parameters and the harmonic coefficients:

$$R = R\left(a, e, \Omega, i, \omega, \overline{M}, J_{lm}, K_{lm}\right). \tag{5.34b}$$

The second-order differential equation (5.34) can be transformed into a system of first-order differential equations. They represent the time rates of the orbital parameters as a function of partial derivatives of the perturbing potential (i.e., of the harmonic coefficients) with respect to them. These first-order differential equations are known as Lagrange's perturbation equations (KAULA 1966, p.29, SEEBER 1993, p.75):

$$\left.\begin{aligned}
\frac{da}{dt} &= \frac{2}{na}\frac{\partial R}{\partial \overline{M}} \\[2mm]
\frac{de}{dt} &= \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial \overline{M}} - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial \omega} \\[2mm]
\frac{d\omega}{dt} &= -\frac{\cos i}{na^2 \sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i} + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial e} \\[2mm]
\frac{di}{dt} &= \frac{\cos i}{na^2 \sqrt{1-e^2}\sin i}\frac{\partial R}{\partial \omega} - \frac{1}{na^2 \sqrt{1-e^2}\sin i}\frac{\partial R}{\partial \Omega} \\[2mm]
\frac{d\Omega}{dt} &= \frac{1}{na^2 e\sqrt{1-e^2}\sin i}\frac{\partial R}{\partial i} \\[2mm]
\frac{d\overline{M}}{dt} &= n - \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial e} - \frac{2}{na}\frac{\partial R}{\partial a}
\end{aligned}\right\}. \tag{5.35}$$

The influence of the *gravitation* of *moon* and *sun* on a satellite can be calculated by corresponding extension of (5.27), which leads to the equation of motion for a four-body problem. As a result, a satellite orbit experiences secular and long-periodic perturbations, which may reach 100 m and more. In addition, *solid earth* and *ocean tides*, cf. [8.3.5], especially affect low-flying satellites. Corresponding corrections are based on the ephemeris of the moon and sun and on earth and ocean tide models.

Air drag is caused by friction of the satellite with atmospheric particles. It is proportional to the velocity of the satellite and depends on atmospheric density and the effective cross-sectional area to mass ratio. With increasing altitude, the air drag decreases rapidly and approaches zero at about 1000 km. Air drag

effects are corrected using high-altitude atmospheric models such as CIRA72 (Cospar International Reference Atmosphere 1972), cf. [5.1.2].

High-altitude satellites are especially affected by *solar radiation pressure* due to incident photons. The resulting perturbations depend on the attitude of the satellite with respect to the sun, the area to mass ratio, the reflectivity, and the solar flux. The earth-reflected solar radiation pressure (albedo) remains significantly smaller than the direct effect. Modeling of the radiation pressure effects is difficult, especially for satellites of complex structure. *Electromagnetic interactions* with the magnetic field of the earth occur in the ionosphere; however, they are small and can be considered by corrections. At the present level of accuracy, relativistic effects also have to be taken into account.

Orbit determinations are based on analytical or numerical methods (BOCCALETTI and PUCACCO 1996/1999). For *analytical solutions*, all acting forces are expressed by rigorous relations and integrated in closed form. A first order approximation is already provided by the solution of (5.35). The position vector at any epoch t is given by the orbital elements at an initial epoch t_0, the parameters of the gravitational field, and other models of disturbing forces:

$$\mathbf{r} = \mathbf{r}\left(a_0, e_0, \Omega_0, i_0, \omega_0, \bar{M}_0; GM, J_{lm}, K_{lm}; \ldots; t\right). \qquad (5.36)$$

For *numerical methods*, all forces are calculated for a particular position of the satellite and used as a starting condition for a step-wise integration. Classical astronomic methods for orbit determination are used, as developed by *Cowell* (integration of the total force) or *Encke* (integration of the difference to an osculating Kepler ellipse). Conventional integration methods based on a polynomial fit to consecutive points (e.g., the *Runge-Kutta* method) provide a rigorous solution.

The analytical method is rather laborious, and difficulties arise at applying it to non-gravitational forces. It is well suited for estimating the effects of perturbing forces on the satellite's orbit and for the planning of satellite missions and projects. Numerical methods are simple and generally applicable. They are used nearly exclusively today.

With precise models for the perturbing forces, orbit predictions can be extended to several revolutions of the satellite. Predicted orbits for individual satellites are published by the responsible agencies, while navigation satellites transmit their own orbital data, cf. [5.2.4], [5.2.5]. Precise orbit determinations are necessary for positioning and satellite altimetry. On the other hand, an orbit analysis can be used for deriving or improving the harmonic coefficients of the gravitational field, cf. [6.6.2].

Under good satellite tracking, and with precise models of both the gravitational field and the non-gravitational forces available, the ephemeris of laser and altimeter satellites can be determined with an accuracy of a few cm for short arcs (fractions of a day) and with dm-accuracy for long arcs (few days). The orbital data of high-altitude navigation satellites are predicted with an accuracy of a few m, and orbital post-processing using tracking data of dedicated networks reaches sub-dm accuracy, cf. [5.2.5].

5.2.3 Artificial Earth Satellites

Since the launch of Sputnik I (1957), artificial earth satellites have been used for geodetic purposes such as positioning and the determination of the earth's gravity field and rotation parameters. Only a few satellite missions have been designed exclusively for geodetic applications. However, a large number of satellites developed for navigation, remote sensing, and geophysics were and are used extensively in geodesy.

A satellite can be regarded as a moving *target* at high altitude and then used mainly for positioning. Because the satellite's orbit is affected by the gravitational field of the earth, the satellite may also serve as a *sensor* for gravitation. Satellites may reflect incident light only (passive satellites), or they may carry on board subsystems such as transmitters/receivers, different type sensors, clocks, and computers (active satellites). In the latter case, an energy supply is required, and lifetime is rather limited.

The mean *orbital velocity* of a satellite moving in an approximately circular orbit ($r = a$) is given from (5.29) by

$$\bar{v} = a\bar{n} = \left(GM/r\right)^{1/2}. \tag{5.37}$$

For a satellite close to the earth ($h = 1000$ km), we obtain, with $r = R + h = 7370$ km, a velocity of 7.4 km/s. Kepler's third law yields the period of revolution $U = 2\pi r/\bar{v} = 104$ min.

The intersection of the orbital plane with a non-rotating earth represents a great circle on the earth's surface: *subsatellite track*. The rotation of the earth causes a western displacement of subsequent satellite orbits (Fig. 5.6), with a shift on the equator given by

$$\Delta\lambda = 360° \frac{U}{\text{sidereal day}} = 15° \cdot U\left[\text{h}\right] = 0.25° \cdot U\left[\text{min}\right]. \tag{5.38}$$
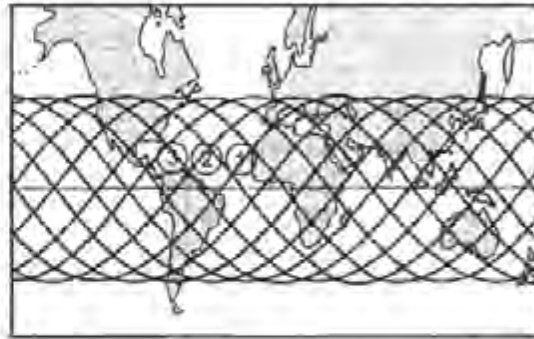
Fig. 5.6. Subsatellite tracks

The latitude range of the subsatellite tracks is determined by the inclination of the satellite.

The following aspects have to be considered during the *design* (choice of orbital parameters) of *satellite missions* for geodetic applications:

For *positioning*, the network geometry of the ground stations and the satellites plays a primary role. Simultaneous direction measurements from two ground stations to a satellite form a plane, and the intersection of planes provides positions. Range measurements utilize the intersection of spheres, whereas range differences, derived from Doppler-frequency shifts, use the intersection of hyperboloids. Satellites at high altitudes are less influenced by gravitational and air drag perturbations and therefore preferred.

Low altitude satellites are required for the determination of the *gravitational field*. This is mainly due to the attenuation factor $(a_e/r)^l$ in the spherical harmonic expansion of the gravitational potential (a_e = semimajor axis of the earth ellipsoid), cf. [3.3.2]. Consequently, the relative errors of the harmonic coefficients increase rapidly at higher degrees l. For $l = 20$, the amplitude of the corresponding structure of wavelength 2000 km is only 5% at a satellite altitude of 1000 km, as compared to the earth's surface. This attenuation is increased by the fact that the harmonic coefficients become smaller with increasing degree and order, cf. [6.6.2]. In order to resolve the gravitational field with a certain degree l at the equator, it follows from (5.38) that the satellite has to perform 21 revolutions/day, or a longer observation time is required. As seen from (5.35), the recovery of the harmonic coefficients strongly depends on the satellite's inclination. Corresponding satellite coverage is needed in order to avoid ill conditioning; too small inclinations and eccentricities must be avoided.

Gravitational field determination can be based on range and range-rate measurements between satellites and ground stations, as well as between satellites. A high field resolution is achieved by gravity gradiometers carried on

board low-flying satellites. Satellite altimetry provides the distance to sea level and thus a high-resolution approximation to the geoid, cf. [3.4.2]. Both an orientation into the vertical and a high orbital accuracy is required in that case.

Non-gravitational perturbing effects can be reduced by a small cross-sectional surface and large mass; a spherical shape offers special advantages. Atmospheric drag and solar radiation pressure may also be compensated by a drag-free system. In such a system, a proof mass is shielded by a shell attached to the satellite. The mass is affected only by gravitation, while surface forces act in addition on the shell. By continuously measuring the position changes between proof mass and shell, a feedback system keeps the satellite centered on the proof mass.

In order to detect *variations with time*, in position (point velocities). and in the gravitational field, long-term observations have to be carried out, preferably by the same satellite system.

Satellites used in geodesy may be equipped with the following techniques, of which combinations are used in many missions:

- *Direction* measurements have been made available by a sun-light reflecting skin (early balloon satellites), by flashing light devices, and by mirror arrays, cf. [5.2.4],
- transmitters/receivers serve for the continuous emission/reception of modulated radio waves used for *distance* and *Doppler* measurements, cf. [5.2.4], [5.2.5],
- retroreflector arrays of fused silica corner cubes reflect laser light pulses and are employed for *laser distance* measurements, cf. [5.2.6],
- vertical distance measurements to the ocean surface are performed by *radar altimeters*, cf. [5.2.7],
- gravitational *gradiometers* measure the gravitational gradient within the body of the satellite, cf. [5.2.8].

Dedicated systems have been developed specifically for precise *orbit determination*. They consist of space, ground, and control segments. The French DORIS (Doppler Orbitography and Radio positioning Integrated by Satellite) system is based on a worldwide network of beacons emitting radio signals (2 GHz) that are received and processed as Doppler frequency shifts on board the satellite. The German PRARE (Precision Range and range Rate Experiment) is a space borne, two-way, dual-frequency microwave system (2.2 and 8.5 GHz), where the double signal travel time is calculated in the space segment. These systems have operated on different satellites since the 1990's, cf. [5.2.7], and are also exploited for positioning of the ground stations. The radial-orbit accuracy achieved a few cm to 1 dm, and velocities are derived to within ± 0.1 to 0.3 mm/s.

GPS receivers are increasingly installed on satellites for precise orbit determination; they deliver similar accuracies to those discussed above (YUNCK and MELBOURNE 1996).

High demands are placed on the determination of *time*. At orbital velocities of several km/s, the time epoch has to be determined to ± 1 μs in order to keep orbital errors less than 1 cm. Distance measurements to satellites require time interval measurements to ± 0.1 ns in order to obtain cm-accuracy. Rubidium or cesium frequency standards, which are tied to UTC by time signals, are capable of this level of accuracy. Quartz oscillators can be used in satellite receivers if an external control is provided, e.g., through the satellite system, cf. [5.2.5].

More details on satellites employed in geodesy are given in the following chapters.

5.2.4 Direction, Range and Range Rate Measurements: Classical Methods

Satellite observations began in 1957. They were based partly on methods developed for the observation of the moon and of balloons. While some of the classical techniques applied up to the 1980's are no longer of importance, other such methods have been developed further and are still used extensively today. Some results of the early satellite missions are still of relevance for the strengthening and orientation of geodetic networks and for the calculation of earth models, cf. [6.8.1].

Direction measurements to satellites prevailed until about 1970 and led for the first time to global and regional three-dimensional networks. They also provided the low-degree harmonic coefficients of the gravitational potential.

For optical direction measurements, an illuminated satellite is photographed on film or glass plates together with the fixed stars.

Balloon satellites reflecting both sunlight (e.g., Pageos, 1966-1972, diameter of 30 m, $i = 87°$, $h = 2800$ to 5600 km) and light flashes from active satellites were used. Ballistic cameras (e.g., Wild BC4) mounted azimuthally were easy to operate and had a large field of view. Equatorially-mounted astronomic cameras could follow the motion of the stars. Using a large focal length, even faint stars could be detected and observed with high precision. Orbital cameras were designed so that they could also follow the motion of a satellite.

Upon developing the photographs and identifying the stars, the satellite and star image points were measured on a precision comparator (± 1 μm). The plate coordinates of the satellite were then transformed to the spatial directions right ascension and declination, whereby the transformation parameters were derived from the known directions to the stars, cf. [2.4.1]. Since the satellite travels in the atmosphere, astronomic refraction reduction as applied to the stars, cf. [5.3.3], had to be reduced by the satellite refraction (Fig. 5.28), with

$$\Delta z_{sat} = 0.48'' \frac{tan\,z}{h[1000\ km]}. \tag{5.39}$$

The *accuracy* of the direction measurements was ±0.2" to 2". An increase in accuracy was not possible mainly due to scintillation effects and comparator measuring errors.

The Japanese satellite EGS (also Ajisai, h = 1500 km) is a recent example of direction measurements. The satellite is equipped with both mirror and laser reflectors. By rotation of the satellite about its axis, incident sunlight is reflected periodically with two flashes per second.

Microwave distance-measurements were employed in the 1960's and still play a fundamental role today.

The *Secor* (Sequential Collation of Range) system used modulated microwaves (two frequencies) emitted from ground stations and re-transmitted from transponders on board the satellites. Distances were derived from signal's travel time, using the phase comparison method, cf. [5.5.2]. Using this method and employing spatial trilateration, isolated geodetic networks were connected to a global reference system. The precision obtained was a few m, although systematic errors up to several 10 m also occurred.

From the 1970's to the 1980's, *Doppler positioning* became an operational tool for establishing or improving geodetic networks. By orbital analysis, it also delivered improved geopotential models and earth rotation parameters.
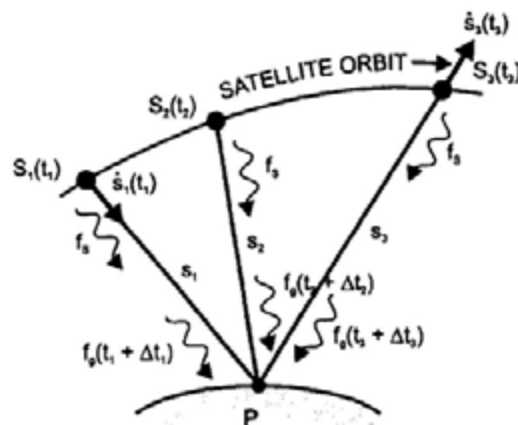
Fig. 5.7. Satellite Doppler positioning

With Doppler measurements, a transmitter aboard a satellite S continuously emits a stable frequency f_s (Fig. 5.7). A signal is received at the ground station with the frequency f_g and a time shift Δt with respect to the transmission time t. f_g is shifted against f_s due to the relative velocity $\dot{s} = ds/dt$ between the satellite and the observer (Doppler effect):

$$f_g - f_s = -\frac{f_s}{c}\dot{s}\,.\tag{5.40}$$

The Doppler frequency shift is proportional to $\dot{s}$; a reversal in sign occurs at the time of the closest approach of the satellite to the observer ($\dot{s} = 0$). In principle, a range difference (range rate) can be determined from (5.40) by integration over time. In practice, f_g is compared with a constant reference frequency f_0 generated within the Doppler receiver, with $f_0 \approx f_s$. Integration over a time interval yields the Doppler count

$$N_{ij} = \int_{t_i+\Delta t_i}^{t_j+\Delta t_j} (f_0 - f_g)dt\,.\tag{5.41}$$

With (5.40), we obtain the observation equation

$$N_{ij} = (f_0 - f_s)(t_j - t_i) + \frac{f_0}{c}(s_j - s_i)\,,\tag{5.42}$$

which provides the range rate $s_j - s_i$ from the Doppler counts.

As with any microwave technique, Doppler measurements do not depend on weather conditions, and they allow large amounts of data to be accumulated within short time intervals. Today, Doppler measurements are used with several satellite missions and with the DORIS positioning system, cf. [5.2.3].

The *Navy Navigation Satellite System* (NNSS) or *Transit System* was an important application of Doppler measurements (KOUBA 1983, ANDERLE 1986). Developed as a navigation system for the U.S. Navy, it was opened to civil use early on and operated successfully between 1964 and 1996. Positioning was based on 4 to 7 Transit satellites in orbit ($h = 1100$ km, i $\approx 90°$), which continuously transmitted frequencies of 150 and 400 MHz. With the orbital planes of the satellites evenly distributed in longitude, and due to earth rotation, a satellite became visible at least every two hours along the equator. The satellite orbits were determined by Doppler measurements from four tracking stations located in the U.S.A. Initially they referred to the World Geodetic System 1972 (WGS72) and later to WGS84, cf. [5.2.5]. The satellites transmitted their "broadcast ephemeris" at two-minute intervals, together with UTC time signals, with an accuracy of ±10 to 20 m. "Precise ephemeris" (± 1 to 2 m) were later made available to authorized users. A number of Doppler receivers were built as portable units for geodetic purposes. These contained a reference oscillator, a microprocessor, a data-recording unit, the antenna, and an energy supply.

The influence of ionospheric refraction was practically eliminated by the use of two frequencies, cf. [5.1.3], and tropospheric refraction was computed using atmospheric models and observed meteorological data, cf. [5.1.2]. The positional error of ± 10 to 30 m, obtained at the observation of a single satellite pass, could be reduced to ± 2 to 5 m (broadcast ephemeris) with 30 to 50

passes and to ± 0.5 to 1 m using precise ephemeris. Relative positioning employed simultaneous observations on two or more stations, which strongly reduced orbital and refraction errors and led to baseline accuracies of ± 0.2 to 0.5 m (SEEBER et al. 1982).

With the GPS-system becoming operational, the NNSS system was no longer maintained.

5.2.5 Global Positioning System (GPS)

The NAVSTAR/GPS (Navigation System with Time and Range/ Global Positioning System) is a radio navigation system based on satellites. It is operated by the U.S. Department of Defense (DOD), which started development of the system in 1973. The first GPS satellites were launched in 1978, and the system became fully operational in 1993. GPS provides real time navigation and positioning by one-way microwave distance measurements between the satellites and the GPS receivers. Early on, the use of GPS for geodetic applications was investigated (BOSSLER et al. 1980), and GPS positioning is now extensively employed in geodesy at all scales, including the determination of variations with time. Several textbooks on GPS are available, e.g., WELLS (1986), LEICK (1995), HOFMANN-WELLENHOF et al. (1997), see also SEEBER (1993), and for the actual state, TEUNISSEN and KLEUSBERG (1998a), MCDONALD (1999).

The *basic idea* of GPS is to have at least four satellites above the horizon available 24 hours everywhere on the earth. In principle, the three-dimensional coordinates of the receiver's antenna can be derived from three observed distances. The computation is based on the known ephemeris of the satellites and the intersection of spherical shells. As the clocks of the satellite and the receiver are not synchronized, a fourth distance measurement is necessary in order to determine the clock synchronization error. Therefore, the original distances derived from the travel time of a signal are called *pseudoranges* (Fig. 5.8).

We distinguish between the space, the control, and the user segment of GPS.

The *space segment* consists of 21 *satellites* (plus 3 additional spares) arranged in six nearly circular orbits ($i = 55°$, 12 h period of revolution) at an altitude of about 20 200 km (Fig. 5.9, Fig. 5.10). Due to the limited lifetime of a satellite (10 years on the average), a regular replacement is scheduled in blocks, which are intended to correspond to improvements in satellite technology (block II/IIA consisted of 28 satellites launched between 1989 and 1998).

Atomic clocks (2 rubidium and 2 cesium clocks per satellite) provide a high-precision frequency standard, with 10^{-12} to 10^{-13} relative long-term stability, cf. [2.2.1]. They produce the fundamental frequency of 10.23 MHz. By
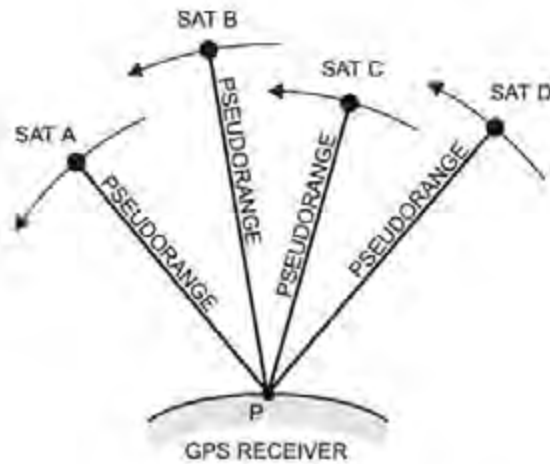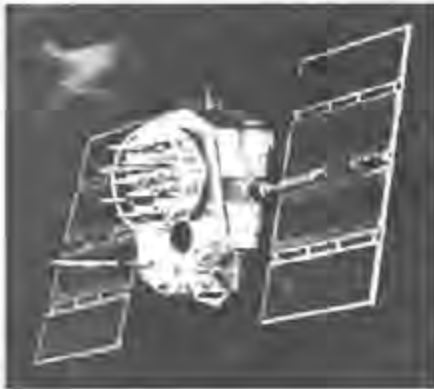
Fig. 5.8 GPS positioning

Fig. 5.9. GPS (Global Positioning System) satellite, courtesy The Aerospace Corporation, El Segundo, CA, U.S.A.

Fig. 5.10. GPS orbit constellation

multiplication, the L1 (1575.42 MHz $\triangleq$ 19.05 cm wavelength) and the L2 (1227.60 MHz $\triangleq$ 24.45 cm) *carrier waves* are derived and continuously emitted. L1 and L2 serve as carriers for two code modulations and for a data signal (navigation message). The codes are given as binary signals (+1 and -1 sequence) in a pseudo-random noise (PRN) form (Fig. 5.11). The *C/A-code* (coarse/acquisition code) is modulated on L1 only, with a frequency of 1.023 MHz ($\triangleq$ 293 m wavelength) and a repetition rate of 1 ms. The *P-code* (precise code) is modulated on L1 and L2 and has a frequency of 10.23 MHz ($\triangleq$ 29.3 m wavelength) and a repetition sequence of 266 days. It primarily serves for
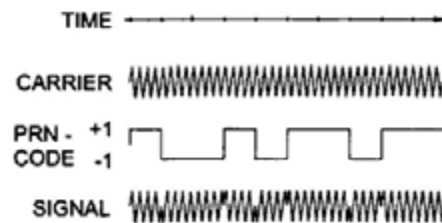
TIME

CARRIER

PRN - +1
CODE -1

SIGNAL

Fig. 5.11. GPS signals

precise navigation and is reserved to authorized users. The *navigation message* is transmitted on L1 and L2. It contains the satellite's ephemeris (Keplerian elements and certain time derivatives and orbital correction terms), a clock correction with respect to GPS-time, the coefficients of a ionospheric refraction model, and information on the status of the GPS system. The broadcast ephemeris are accurate to a few m and approach sub-meter accuracy.

With GPS being fully operational, an *accuracy deterioration* for civil users has been introduced by DOD. The *Precise Positioning Service* (PPS) provides all GPS signals, and thus the full accuracy of the system ($\pm$ 10 to 20 m or better) to authorized users (mainly U.S. military). The *Standard Positioning Service* (SPS) only delivers the C/A-code, while the P-code is encrypted by Anti-Spoofing (AS), which results in the non-accessible Y-code.

Between 1990 and 2000, the fundamental frequency of the satellite clocks was destabilized, and the ephemeris data were manipulated by *Selective Availability* (SA). This introduced errors in the pseudoranges, which degraded the absolute accuracy to about 100 m in horizontal position and 150 m in height. For geodetic purposes, these effects were largely eliminated by differential techniques. The suspension of SA led to an accuracy improvement in positioning and time transfer, reaching the accuracy available for authorized users.

The GPS *control segment* is responsible for maintaining the operation of the GPS satellites. It consists of the master control station (Colorado Springs, CO, U.S.A.) and five globally distributed monitoring stations. The stations are equipped with cesium standards and GPS receivers. They continuously measure pseudoranges to all satellites and transfer the results to the master station. After computation of the satellite orbits and the clock correction, the (extrapolated) broadcast ephemeris and GPS time are transmitted to the satellites for storage and retransmission by three ground antennas collocated with monitoring stations. This operational control system is supplemented by a number of additional monitoring stations (in 1999: 7 stations) operated by the U.S. National Imagery and Mapping Agency (NIMA).

The GPS ephemerides refer to an earth-fixed system, as defined by the coordinates of the monitoring stations: *World Geodetic System* (WGS).

WGS has been developed by DOD since the end of the 1950's (realizations WGS 60, WGS 72, and WGS 84). The system is intended to serve for mapping, charting, positioning and navigation, following international standards for geodetic reference systems. The defining parameters of the latest version of WGS 84 are as follows (SLATER and MALYS 1998, NIMA 2000):

- semimajor axis $a = 6378137\,\text{m}$,
- reciprocal flattening $1/f = 298.257223563$,
- geocentric gravitational constant $GM = 398600.4418 \times 10^9\ \text{m}^3\text{s}^{-2}$, which includes the atmospheric part $GM_{atm} = 0.35 \times 10^9\ \text{m}^3\text{s}^{-2}$,
- angular velocity of the earth's rotation $\omega = 7.292115 \times 10^{-5}\ \text{rad}\,\text{s}^{-1}$.

WGS 84 thus practically coincides with the Geodetic Reference System 1980, cf. [4.3]. The associated gravity field is given by the global geopotential model EGM 96, complete to degree and order 360, cf. [6.6.3].

The coordinates of the monitoring stations are given for the epoch 1997.0, taking earth tides (tide-free system), cf. [3.4.1], and plate tectonic motions into account. The accuracy of the WGS 84 coordinates is ±0.05 m, which is also the level of agreement between WGS 84 and the International Terrestrial Reference Frame (ITRF), cf. [2.5.3].

GPS-time (unit SI second) is defined by the cesium clocks of the control segment stations. It agreed with UTC on January 5, 1980 and differs from it by a nearly integer number of seconds. This difference is increasing because no "leap seconds" are introduced to the GPS time scale, cf. [2.2.2]. The actual difference between GPS-time and UTC is part of the GPS navigation message (± 0.1 μs). GPS is a very efficient system of *time dissemination*. A clock comparison is possible with an accuracy of ± 100 ns and ± 10 ns and better with differential techniques. Special GPS-time receivers are available as C/A one-channel instruments and operate in an automatic mode (LARSON and LEVINE 1999).

A similar satellite navigation system has been developed in the former Soviet Union since the 1970's. GLONASS (Global Navigation Satellite System) also operates as a one-way ranging system. It has been operational since 1996 (LANGLEY 1997a, ZARRAO et al. 1998). Its space segment comprises 24 (including 3 spares) satellites, arranged at a regular spacing of 45° in three nearly circular orbits ($i = 64.8°$, $h = 19\,100$ km, revolution time about 11.2 h). The orbits are spaced 120° apart from each other, with five satellites always being visible. The satellites transmit on two frequency bands (1602-1615 MHz, 1246-1256 MHz) with *different* frequencies for each satellite (contrary to GPS). On the other hand, the C/A- and P-code modulations are the same for all satellites. There is no degradation of the GLONASS signals. The satellites are equipped with laser retroreflectors for laser tracking. The control segment consists of a master control station near Moscow and secondary stations distributed over the territory of the former Soviet Union.

The GLONASS results refer to the reference system PZ90, which is based on the Soviet geodetic reference system SGS85. Transformation parameters to WGS84 are known with an accuracy of a few m (MISRA and ABBOT 1994). GLONASS uses its own time system (UTC + 3 hours) which is synchronized to UTC within 1 μs by the use of leap-seconds.

A European navigation satellite system called GALILEO has been discussed since 1999. It may employ four carrier frequencies, and the intended on-line positional accuracy is 10 m (95%), once every second (DIVIS 1999).

The *user segment* is composed of the many *GPS receivers* operating in navigation, geodesy, and surveying. The main components of a receiver include the antenna, the receiver electronics, the microprocessor, the oscillator, the memory, the user interface, and a power supply. An additional telemetry unit can be used for data transfer between different receivers (differential mode).

The signals transmitted from the satellites are received and amplified by the antenna. After identification, the signals are processed to pseudoranges in the channels of the receiver electronics. One channel is generally responsible for the tracking of one satellite. Hence, a minimum of four channels is necessary to determine position and time. Modern receivers contain 12 or more channels for each frequency (multi-channel technique). The microprocessor controls the operation of the receiver and calculates the three-dimensional position of the antenna in WGS84 and the velocity and azimuth of moving objects. A quartz oscillator is used for generating the reference frequency, which is approximately synchronized with GPS-time. All data (pseudoranges, phases, time, navigation message) are stored in receiver memory for post-processing, which is typical for multi-station observation sessions employed in geodesy, cf. [6.2.1]. The user interface includes a keyboard and a display, which provides a communication link between the user and receiver. Power is provided by internal, rechargeable nickel-cadmium batteries.

Geodetic GPS receivers deliver high accuracies in the static and kinematic mode (BOUCHER 1987, LANGLEY 1997b). This is achieved by dual-frequency (L1 and L2) multi-channel instruments with the P-code and the full carrier wave available on L1 and L2. Other characteristics include low receiver-noise in code and carrier phase, a high data rate (> 1 Hz), and a large memory capacity. The antenna phase center should be stable and protected against multipath effects.

The Macrometer (1982) was the first GPS receiver for geodetic applications (code-free, single frequency, 6 parallel channels). The Texas Instruments TI4100 (1984) provided all geodetically relevant observables (P and C/A-code pseudoranges, carrier phases on L1 and L2) in a multiplex channel technique (SEEBER et al. 1985). The efficiency of geodetic receivers is discussed in the literature (e.g., COORS et al. 1998). Fig. 5.12 to 5.14 show some examples. Combined GPS/GLONASS receivers are also available (Fig. 5.15). They are able to track any combination of GPS and GLONASS satellites (12 to 21 will be visible with fully deployed systems). This will

Fig. 5.12. GPS continuously operating reference station (CORS) with receiver (GPS total station 4700) and choke ring antenna, courtesy Trimble Navigation Ltd., Sunnyvale, CA, U.S.A.

Fig. 5.13. Geodetic GPS dual-frequency receiver (GPS total station SR530), with terminal TR500 and telemetry antenna, courtesy Leica Geosystems AG, Heerbrugg, Switzerland

Fig. 5.14. Geodetic GPS dual-frequency receiver (GePoS Experience) with antenna, courtesy Zeiss/Spectra Precision AB, Danderyd, Sweden

Fig. 5.15. GPS/GLONASS receiver with antenna, courtesy JAVAD Positioning Systems/TOPCON, Paramus, NJ, U.S.A

improve the satellites- receiver geometry and the availability of satellites, as well as the integrity of the solution, and lead to a higher accuracy for absolute positioning with broadcast ephemeris.

A distinction is made between code and carrier phase measurements for GPS-positioning.

Code measurements employ the travel time Δt of a signal between the satellite and the antenna of the receiver. The time difference is determined by cross-correlating an arriving code sequence with a code copy generated in the receiver. Multiplication of Δt with the velocity of light c gives the distance between the satellite and the antenna, cf. [5.1.1]. Considering the clock synchronization error δt, the *observation equation* for the *pseudoranges R* reads as

$$R = c\,\Delta t = s + c\,\delta t .\qquad\qquad (5.43a)$$

The distance is given by

$$s = \left(\left(X_s - X_p \right)^2 + \left(Y_s - Y_p \right)^2 + \left(Z_s - Z_p \right)^2 \right)^{1/2} ,\qquad (5.43b)$$

where X_s, Y_s, Z_s and X_p, Y_p, Z_p are the geocentric coordinates of the satellite and the ground stations respectively. Using the transmitted broadcast ephemeris, the receiver clock correction and the coordinates of the ground station can be derived by simultaneous measurements to four satellites. Equation (5.43) presupposes that atmospheric refraction effects are taken into account by proper reductions (see below).

The measurement noise governs the accuracy limit for this method. Accuracy depends on the wavelength and is at the m-level for C/A-code and at the dm-level for P-code. If used as a standard method in navigation, a positional accuracy of about 10 m is achieved.

Pseudorange differences can be derived from integrated *Doppler frequency shifts* (Doppler counts) of the carrier frequency according to (5.41) and (5.42). The differences are used for the determination of velocity in navigation. Doppler solutions also play a role in the removal of ambiguities, which occur with carrier phase measurements. The method is not suitable for real-time positioning due to the long observation time required.

Relative (differential) GPS methods determine differences of coordinates between two or more stations. This is accomplished either by simultaneous observations on the stations, including at least one with known coordinates, or by using correction data provided by permanently operating reference stations. Corresponding DGPS-Services for navigation deliver range corrections derived from the differences between observed and calculated pseudoranges at reference stations. Accuracies of a half to a few m are achieved over distances between about 50 and several 100 km.

Geodesy and surveying require accuracies of at least two orders of magnitude better than that required for navigation. This is achieved by *carrier phase*

measurements. Due to the shorter wavelength of carrier phases, the measurement noise is only in the mm to sub-mm-range (BEUTLER et al. 1987).

The carrier phase is detected by comparing the received carrier signal with the reference frequency generated in the receiver after subtraction of the code. In order to reconstruct L2 under AS conditions (P-code encryption), different techniques have been developed, such as squaring of L2 (eliminates the code signal) and cross-correlation of L1 and L2. The measured phase difference

$$\Delta\varphi = \varphi_c - \varphi_0 \tag{5.44}$$

(φ_c, φ_0 = phase of the carrier and reference waves respectively) is related to the distance s by

$$\Delta\varphi = \frac{2\pi}{\lambda}(s - N\lambda + c\delta t), \tag{5.45}$$

which is well known from terrestrial distance measurements, cf. [5.5.2]. N is an integer number of complete carrier cycles within the range s, and δt is the receiver clock synchronization error. The *ambiguity* introduced by N poses a primary problem for the evaluation of (5.45). Among the algorithms available for ambiguity determination, we have the inclusion of ambiguity-free Doppler solutions, the combination of code and carrier phases, and statistical search methods applied to combinations of L1 and L2. Difficulties arise when the phase lock is lost due to signal obstruction. Such sudden jumps of the carrier phase by an integer number of cycles are called *cycle slips*. They are either removed during preprocessing or taken into account by introducing an additional ambiguity for the affected pseudorange.

The GPS signals experience path delays while traveling through the atmosphere,. *Ionospheric refraction,* cf. [5.1.3], causes range errors at the m to 100 m order of magnitude (KLOBUCHAR 1991). Using the satellite message, a reduction can be applied to the travel time when only a single-frequency receiver is available. Two-frequency receivers allow application of (5.23) to the frequencies f_1 and f_2, which leads to the distance (code measurement)

$$s = \frac{s_1 f_1^2 - s_2 f_2^2}{f_1^2 - f_2^2}. \tag{5.46}$$

In (5.46) the ionospheric refraction is eliminated, and s_1 and s_2 are the observed distances on L1 and L2 respectively. A corresponding equation can be derived for carrier phase observations. *Tropospheric refraction* may cause delays on the order of one m (zenith direction) to 10 m (close to the horizon).

Refraction is accounted for by tropospheric models and observed surface weather data, cf. [5.1.2]. The "wet" component remains a critical part of these reductions; an improved determination is possible by water vapor radiometers, which measure the water vapor content along the path.

Multipath effects result from signal reflection near the antenna. They affect code and carrier measurements and may produce errors on the order of cm to dm. A reduction of these effects is possible by a proper design of the antenna and by careful site selection. The antenna's phase center (close to the geometric center) varies in dependence on the satellite's elevation and azimuth; this must be taken into account by calibration (MENGE et al. 1998).

The *error budget* of GPS is composed of orbital and satellite clock errors, residual refraction effects, and receiver errors. Orbital errors are at the order of a few m (Standard Positioning Service).

High-quality GPS data and orbits are available for post-processing through the International GPS Service (IGS), which operates under the auspices of IAG, cf. [7.3.1]. IGS provides, among other things, a precise and a "rapid" GPS ephemeris. The precise ephemeris is available after 7 days and has a precision of about ±3...5 cm. The "rapid" ephemeris is available after 24 hours and is only slightly less precise. Satellite clock parameters are given with a precision of ±1 ns. The accuracy of these products should be close to the precision.

Clock errors are about 1 m. The residual errors of ionospheric refraction reach only a few cm with two-frequency receivers, while tropospheric model errors may attain one dm. The receiver noise is roughly about 0.1% of the signal's wavelength, which corresponds to a dm-order of magnitude with the C/A- code and 1 to 2 mm with the P-code (LANGLEY 1997b). Phase center variations of the antenna may reach a few cm.

The *accuracy* of GPS positioning also depends on the geometric constellation of the satellites with respect to the receivers and on the duration of the observation time. A longer observation time increases the accuracy, especially for long baselines and for the height component. The accuracy of the kinematic mode (moving GPS receiver) is lower than that of the static mode.

The accuracy of an observed pseudorange can be expressed by its standard deviation (also user equivalent range error). The strength of the satellite geometry is characterized by a quantity called "Positional Dilution of Precision" (PDOP). It is defined as the ratio between the standard deviation of a position derived from a certain satellite constellation and the standard deviation of an observed pseudorange (LANGLEY 1999). The numerator of this quotient follows from the trace of the coordinates covariance matrix, which depends on the design of the network, e.g., SEEBER (1993, p.292). PDOP values can be calculated in advance and serve for the planning of observations and for rapid information on expected positioning quality. For instance, a PDOP value of 3 means that the accuracy of positioning is three times worse than the accuracy of the

pseudorange observation. If separated into the horizontal and the vertical components, it turns out that the determination of heights is less accurate than horizontal positioning by a factor of about two. This results from the fact that all observed satellites are above the receiver but distributed over the total horizon.

Relative positioning (differential GPS) significantly reduces the errors that occur in absolute positioning, especially orbital and refraction errors. This is especially valid for shorter baselines, due to the correlation of these errors at the two ground stations. For orbital errors, a (pessimistic) rule of thumb allows an estimate of the error to be expected in a baseline (length b) from the orbital error dr:

$$\frac{db}{b} = \frac{dr}{s},$$
(5.47)

where s is the distance between the satellite and the receiver (maximum 25 000 km). Hence, if an accuracy of 1 cm is required for the baseline, the orbital error should not exceed 2.5 m at b = 100 km and 0.25 m at b = 1000 km. When the precise ephemerides from the IGS are used, orbital errors no longer play a large role. The obtainable accuracy of relative positioning with good PDOP conditions and in the static mode is about

$$\pm\left(3...5\,mm + 0.5...1\times10^{-6}\,s\right)$$

for the horizontal coordinates. In the kinematic mode, this reduces to

$$\pm\left(10...20\,mm + 1...2\times10^{-6}\,s\right).$$

Height determination is less accurate by about a factor of two.

The Global Positioning System has drastically changed surveying methods in geodesy, navigation, and other applications. This is mainly due to the high accuracy achieved with static and kinematic positioning, real-time evaluation, and operational flexibility. Direct visibility between the ground stations is not necessary; only visibility to the satellites is required. The system is weather independent and usable day and night. The use of GPS is still increasing and is strongly supported by global and regional GPS services, cf. [7.3.1], [7.3.2].

5.2.6 Laser Distance Measurements

Laser distance measurements may be made to satellites equipped with corner cube reflectors: *Satellite Laser Ranging* (SLR). This method offers high accuracy due to the favorable propagation of laser light in the atmosphere. It

also offers a low-cost, long-lifetime space segment. On the other hand, SLR depends on weather conditions and requires a considerable operational effort at the ground segment.

At the ground station, laser pulses are emitted at epoch t, reflected at the satellite, and received again at epoch $t + \Delta t$. If refraction effects are taken into account by corresponding reductions, the distance is obtained by

$$s = \frac{c}{2}\Delta t .$$
(5.48)

Nowadays, travel time of the photons can be measured with an accuracy of ±10 to 100 ps, which corresponds to a ranging accuracy of ±1.5 to 15 mm (BENDER 1992).

Many *satellites* (about 30 in 1999) are equipped with laser reflector arrays for precise orbit determination, e.g., satellites carrying radar altimeters for ocean height determination, cf. [5.2.7]. Dedicated SLR missions for positioning and geodynamic research include the satellites Starlette (France, 1975, h = 800 to 1100 km, i = 50°), Lageos 1 and 2 (U.S.A., 1976/1992, h = 5900 km, i = 110°/52°, COHEN et al. 1985), Ajisai (Japan, 1986, h = 1500 km, i =50°), and Etalon 1 and 2 (USSR, 1989, h ≈ 19 000 km, i = 65°). These satellites are spherical in shape (diameter 0.2 to 2 m) and possess a favorable surface-to-mass ratio (Fig. 5.16). The GLONASS satellites and a few GPS satellites also carry laser retroreflectors on board.
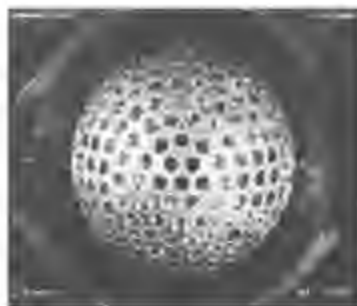

Fig. 5.16. Laser satellite LAGEOS, courtesy National Aeronautics and Space Administration (NASA)

Laser distance *measuring systems* have developed rapidly since the 1960's. They consist of the laser unit (Nd:Yag-laser = Yttrium-aluminum garnet crystal doped with neodymium ions), the transmitting and receiving optics (telescopes), and the receiver electronics (secondary-electron photomultiplier). The mechanical mounting provides an automatic tracking of the satellite's pre-

calculated orbit. Due to atmospheric disturbances and reflection, only a few laser pulses (photons) return. The travel time is measured by a time-interval counter and an atomic clock, regularly compared with UTC. A process computer controls the complete measurement and registration.

About 50 *laser-satellite-systems* are operating worldwide (as of 1999), either in the stationary or mobile mode. The Wettzell laser ranging system employs a Nd:Yag laser (532 nm) and a 75 cm telescope. It operates with high-energy short pulses (pulse length 180 ps, pulse energy 100 mJ) at a pulse repetition rate of 1 to 10 Hz in the single-shot mode. Visible and /or infrared light is used, allowing a day-and-night operation to satellites at altitudes between 800 and 40 000 km and to the earth's moon (DASSING et al. 1992), Fig. 5.17. Further developments are directed to reduce the pulse length to some 10 ps and to reduce the pulse energy. The emission of short-pulse trains increases the probability of single photon detection.

Fig. 5.17. 75 cm telescope, Wettzell Laser Ranging System (WLRS), Fundamentalstation Wettzell, Germany, courtesy Bundesamt für Kartographie und Geodäsie (BKG), Frankfurt a.M., Germany

Mobile systems have been developed in the U.S.A. (NASA) and in Germany/Netherlands, among others (SILVERBERG 1978). Operating with low energy and single-photon detection, these systems are employed mainly for the investigation of recent crustal movements, cf. [8.3.3]. A transportable, integrated geodetic-observatory has been developed by BKG (Bundesamt für Kartographie und Geodäsie, Germany). In addition to the laser unit, it includes a VLBI module and GPS unit. It is employed for strengthening fundamental reference networks, especially in the southern hemisphere.

The *accuracy* of laser distance measurements depends on the pulse length, the stability of the photomultiplier, and the time resolution. Atmospheric refraction effects may reach 2 m at the zenith and 10 m at an elevation of 10°; they are corrected with standard atmospheric models. Depending on the satellite's altitude, some 100 to 1000 distances are measured during one passage, with a

single shot precision of ±5 to 10 mm. By compressing the data to "normal points" (average over 30 s to 120 s at Lageos), sub-cm precision is achieved.

Laser reflectors have been placed on the *moon* by the Apollo 11 (1969), 14, and 15 missions and by Luna 21/ Lunokhod 2. These reflectors provide targets for *lunar laser ranging* (BENDER et al. 1973). Pulsed lasers with a tightly bundled beam and a powerful telescope are necessary in order to recapture the weak returning signal (single photon technique). The tracking system must provide a 2 arcsec pointing accuracy.

Observations to the moon have been carried out regularly since 1969 by the McDonald Observatory, University of Texas, and by systems operating in Hawaii and Grasse/France. Results have also been provided by laser stations in Australia and Germany (Wettzell). The accuracy of these measurements is about 1 cm. They especially contribute to investigations of the dynamics of the earth-moon system and on terrestrial geodynamic processes (SMITH et al. 2000).

5.2.7 Satellite Altimetry

Satellite altimetry is based on a satellite-borne *radar altimeter* that transmits pulses in the vertical direction to the earth's surface. The ocean surface reflects the pulses perpendicularly, and the measurement of the travel time Δt furnishes the height of the satellite above the instantaneous sea surface (Fig. 5.18):

$$a = \frac{c}{2} \Delta t .$$

(5.49a)

(The denotation a should not be confused with the abbreviation for the semimajor axes of the satellite's orbit and the earth ellipsoid). A proper reduction of atmospheric refraction effects is presupposed in (5.49).
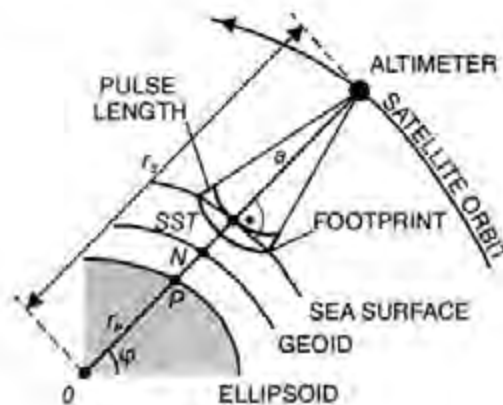

Fig. 5.18. Satellite altimetry

In spherical approximation, this observation can be expressed as

$$a = r_s - r_p - (N + \text{SST}),\qquad\qquad (5.49\text{b})$$

where r_s and r_p are the geocentric distances to the satellite and to the subsatellite point P on the ellipsoid; N is the geoid height and SST the height of the surface topography. Satellite tracking provides r_s and positioning gives r_p. Altimetry thus delivers information on the geoid and on sea surface topography. Ellipsoidal formulas are given by GOPALAPILLAI (1974).

Radar altimeters operate in the 14 GHz frequency range with short (a few ns) pulses and a high-pulse frequency (e.g., 100 pulses/s). The effects of beam divergence and finite pulse length result in measurements that refer to a "mean" sea surface within a circular "footprint" (few km diameter); short-wavelength features of the ocean (waves) are thereby smoothed out. For example, by averaging the measurements over 1 s, the along-track resolution is about 7 km. Satellite altimetry missions are designed to provide either an exact repetition of ground tracks (days to weeks) or a dense pattern of profiles. The different modes are achieved by orbital maneuvers (KNUDSEN 1993). The latter is for determination of the altimetric geoid and the former for investigation of ocean variability (Fig. 5.19).
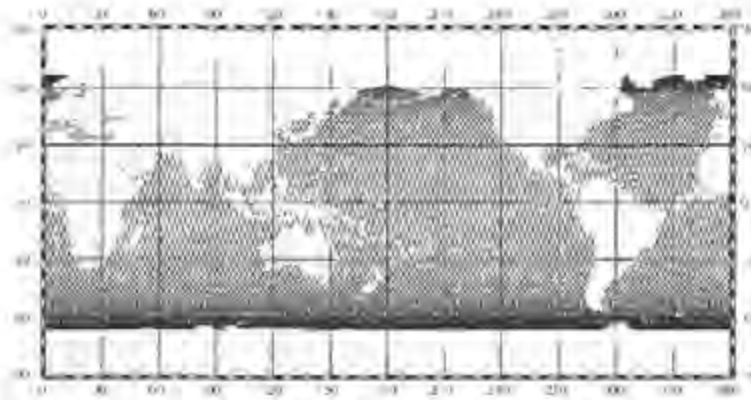

Fig. 5.19. TOPEX/Poseidon subsatellite tracks covered within the 9.9 days repeat cycle, from BOSCH (2001)

The first global survey with a radar altimeter was accomplished by the Geos-3 satellite (U.S.A., 1975-1978). The oceanographic satellites SEASAT (1978) and GEOSAT (U.S.Navy, 1985-1989) carried improved altimeter systems and operated in heights close to 800 km with 108° inclination and repetition rates of 3 and 17 days (TAPLEY et al. 1982, McADOO and SANDWELL 1988). The European Remote Sensing Satellites ERS-1 (1991) and ERS-2 (1995) operate at similar heights

with 98° inclination (Fig. 5.20). Repetition rates are 35 and 168 days respectively, and ground track distances at the equator are 80 km and 8 km respectively for geodetic missions (KOCH 1989, JGR 1998). The NASA/CNES (French space agency) TOPEX/Poseidon satellite (1992, 5.3 and 13.6 GHz) is placed in a circular orbit at an altitude of 1340 km and an inclination of 63°. Repetition time is 10 days, and the equatorial ground track interval 316 km (FU et al. 1994, CHENEY et al. 1995), see Fig. 5.21.

Fig. 5.20. European Remote
Sensing (altimeter) Satellite ERS,
courtesy European Space Agency
(ESA)

Fig. 5.21. TOPEX/Poseideon
(altimeter) satellite, courtesy
JPL/NASA, Pasadena, CA, U.S.A.

High radial-orbit accuracy is required for satellite altimetry. Consequently, the satellites are equipped with laser reflectors, with some having active systems such as Doppler transmitters, C-band radar transponders, and GPS. The space-borne orbit determination systems DORIS and PRARE have been installed on the TOPEX/Poseidon and ERS-2 respectively (TAPLEY et al. 1995, ANDERSEN et al. 1998), cf. [5.2.3]. Further orbital improvements have been achieved by "tailored" gravitational field models developed for a dedicated altimeter mission (TAPLEY et al. 1996). The orbital error thus has been reduced to about ±0.5 m for the GEOS-3 mission, to ±3 cm for TOPEX/Poseidon, and to a range of ±5 to 10 cm for ERS-1. For the evaluation of a region sampled multiple times with high spatial-resolution altimeter profiles, the track crossover discrepancies are adjusted by minimum conditions with time-dependent, low-order polynomial models (e.g., shift and inclination) applied on the individual tracks (van GYSEN and COLEMAN 1997). Residual orbit and other long-term errors are further reduced in this way, cf. [6.6.3].

The altimeter instrumental noise is now less than 2 to 3 cm (TOPEX/Poseidon, ERS). Corrections have to be applied for systematic instrumental (bias and drift) and sea-state effects as well as for tropospheric and ionospheric path delay (modeling, use of two frequencies for TOPEX/Poseidon), cf. [5.1.3]. After reduction of the ocean tides and large-scale air pressure effects, the altimetric

results refer to the quasi-stationary sea surface and yield its height with an accuracy of ±5 to 10 cm.

5.2.8 Satellite-to-Satellite Tracking, Satellite Gravity Gradiometry

High-resolution gravity-field determination from space requires low-orbiting satellites and highly sensitive sensors, cf. [5.2.3]. This is achieved by satellite-to-satellite tracking and satellite gravity gradiometry (COLOMBO 1989, SNEEUW and ILK 1997, ILK 2000).

Satellite-to-satellite tracking (SST) employs microwave systems for measuring range rates between two satellites. High-low (one high and one low-flying satellite) and low-low (two low-flying satellites at the same altitude) configurations have been designed (RUMMEL et al. 1978), Fig. 5.22. The basic observables are the range rates (radial velocities) and changes of the range rates, which are due to gravitational and non-gravitational "disturbing" forces, cf. [5.2.2]. The gravitational field parameters (harmonic coefficients) can be derived after proper compensation of the surface forces. In order to achieve a gravity field resolution of 100 km, the orbital altitude of the lower satellite must not exceed a few 100 km. The relative velocity between the satellites has to be determined with an accuracy of ±1 to 10 μm/s, and precise tracking should be guaranteed by high-altitude satellite systems (GPS) and ground stations.
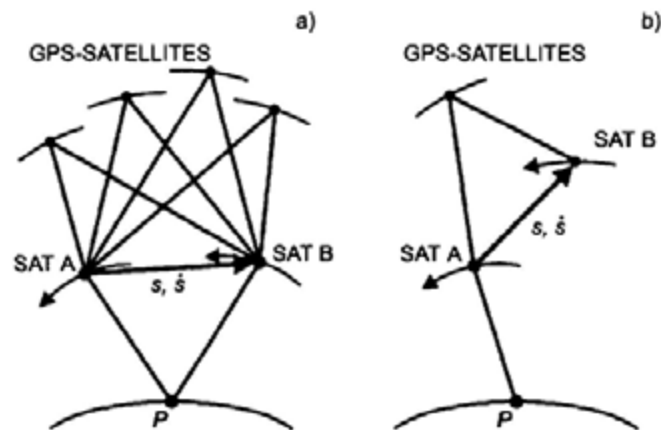

Fig. 5.22. Satellite-to-satellite tracking: a) low-low mode, b) high-low mode

SST experiments started in the 1970's, for example, between GEOS-3 and the geostationary satellite ATS-6 (nearly circular equatorial orbit, h ≈ 36 000 km). The CHAMP (Challenging Mini-Satellite Payload for Geophysical Research and Application) satellite (Germany, launch in 2000) began at an altitude of 450 km, which will be reduced to 300 km within the mission time of 5 years. It moves in a nearly circular polar-orbit (i=83°). The satellite (dimensions 4 m x 1 m x

1.6 m) carries a GPS receiver (high-low mode tracking), laser reflectors, a three-axes accelerometer ($\pm 10^{-8}$ ms^{-2}, for measuring and later compensation of the surface forces), and magnetometer systems (Fig. 5.23). A gravity field recovery up to degree and order 50 is expected (REIGBER et al. 2000).

Fig. 5.23. Champ satellite, courtesy Geoforschungszentrum (GFZ) Potsdam, Germany

The GRACE (Gravity Recovery and Climate Experiment) mission (U.S.A./Germany, launch scheduled for 2001, mission duration 5 years) will employ two satellites of the CHAMP type in the low-low mode about 200 to 300 km apart at an initial orbit altitude of 450 km (i = 89.5°). The relative range measurements between the satellites will be integrated by GPS tracking. A gravity field resolution up to degree and order 150 to 180 will be achieved, with accuracies of $\pm$ 10 to 30 μms^{-2} for the gravity anomalies and $\pm$ 0.1 m for the geoid.

An essential mission goal is the detection and monitoring of large-scale gravity field variations with time, cf. [8.3.4].

Satellite gravity gradiometry determines the components of the gravitational gradient tensor (second derivatives of the gravitational potential). On the earth's surface, gravity gradiometry has been employed since about 1900 with sensor pairs (accelerometers) sensitive to local changes of the gravity field in a certain direction. By different orientation of the sensors, different components of the gravity gradient can be determined, cf. [5.4.5]. For space-borne applications, the attenuation of the gravity field with height (at a few 100 km height, the off-diagonal elements of the gradient tensor are a few 10^{-9} s^{-2} only) requires a high accuracy for the second derivatives (at the order of 10^{-11} to 10^{-13} s^{-2}), which can be achieved with conventional or superconducting electronics (WELLS 1984, PAIK et al. 1988). High demands are posed on the attitude control of the sensor pairs and on the drift of the accelerometers. Surface forces cancel when the output of an accelerometer pair is differenced (MORITZ 1968c).

ESA is planning a dedicated gradiometry mission GOCE (launch 2004?), with a drag-free satellite (h = 270 km, i = 97°) in a sun-synchronous near circular orbit. Tracking will be performed by GPS, and earth-pointing orientation is provided by star trackers. Accelerometer pairs ($\pm 10^{-12}\,s^{-2}$) will be arranged over three mutually orthogonal directions, with baselines of 70 cm. A mission duration of 12 months should provide a gravity field resolution up to degree and order 300, with accuracies of about $\pm$ 10 μms^{-2} (gravity anomalies) and $\pm$ 1 cm (geoid). The low-degree potential terms will be determined primarily through GPS tracking, while the higher degrees will result from gradiometry (RUMMEL 1986, RUMMEL et al. 2000b).

5.3 Geodetic Astronomy

Classical geodetic astronomy is concerned with the determination of astronomic latitude, longitude, and azimuth from ground-based optical direction measurements to fixed stars, which also requires time determination (MUELLER 1969, SCHÖDLBAUER 2000). Several types of observational instruments are available for this purpose [5.3.1], and different methods of observation have been developed [5.3.2]. A number of reductions are necessary in order to refer the observations to the celestial reference frame [5.3.3].

Geodetic astronomy is based on spherical astronomy (EICHHORN 1974, KOVALEVSKY 1995). Its importance has decreased since the development of efficient satellite positioning and gravimetric methods and is now restricted to more local applications of gravity field (plumb line direction) and azimuth determinations. On the other hand, radio waves emitted from extragalactic sources are used extensively in order to derive base-line vectors between fundamental terrestrial stations and to determine earth-rotation parameters: Very Long Base Line Interferometry [5.3.4].

5.3.1 Optical Observation Instruments

Optical observations to fixed stars are carried out in the local astronomic (horizon) system. The direction to a star is determined by the astronomic azimuth A and the zenith angle z (sometimes the altitude or elevation angle 90° - z is used), cf. [2.6.2]. Due to the movement of the observer with respect to the stars, simultaneous time measurements are required.

Astronomic instruments are either permanently installed in observatories or constructed as transportable devices for field operation.

Observations of highest precision with *stationary* instruments have been utilized by the former International Time Service and International Polar Motion Service. Among other instruments, the photographic zenit tube has been employed. In this case, stars near the zenith are photographed symmetrically with respect to the meridian. The zenith angles and the hour angles can be

determined from the tracks of the stars. The direction of the vertical is established by a mercury pool. Comparable precision ($\pm 0.05''$) was achieved by the Danjon prism astrolabe. Thereby, the time is determined when a star crosses the horizontal circle of a given zenith angle. These astronomic observatory measurements are no longer of significance for geodesy, as celestial and terrestrial reference-frames are now established by other space techniques, cf. [2.3].

The *universal instrument* was employed for field measurements of first-order precision ($\pm 0.1''$ to $0.3''$). It consists of a high-precision theodolite, cf. [5.5.1], of very stable design with a few attachments for astronomic observations.

An angled telescope with a horizontal eyepiece permits observations near the zenith. To eliminate personal errors, the movable thread of the registering micrometer is driven to follow the star so that impulses are generated and recorded at uniform intervals. The suspension level serves to measure the tilt of the horizontal axis. The Horrebow level, mounted at right angles to the horizontal axis, registers any changes in the tilt of the telescope. The Kern DKM3-A and the Wild T4 universal theodolites were used widely (Fig. 5.24).

Fig. 5.24. Universal instrument Kern DKM3-A, courtesy Kern/Leica Geosystems AG, Heerbrugg, Switzerland

Fig. 5.25. ZEISS Ni2 level with prism astrolab, courtesy Carl Zeiss, Oberkochen, Germany

The *prism astrolab* is used for the simultaneous observation of astronomic latitude and longitude. With this device, one measures the transit times of those stars that cross the same small-circle parallel to the horizon (almucantar). The constant zenith distance (usually $\approx 30°$) is realized by a prism placed in front of the telescope, and the direction of the vertical is defined by the surface of a pool of mercury or by a compensator pendulum. Astrolabe attachments are

particularly common. These are mounted either on a theodolite (e.g., the Wild T3 astrolabe with a mercury pool) or on an automatic level (Zeiss Ni2 astrolabe), DEICHL (1975), Fig. 5.25.

Transportable *zenith cameras* are also used for rapid determination of latitude and longitude. Such an instrument consists of a camera oriented in the direction of the plumb line (focal length 300 to 1000 mm, relative aperture ≈ 1:5), which can be rotated around the plumb-line axis in any azimuth. In addition to a timing device, the instrument is also furnished with two levels that are arranged at right angles to each other (Fig. 5.26).

Fig. 5.26. Transportable zenith camera, Institut für Erdmessung, Universität Hannover

The field of stars near the zenith is photographed for two positions of the camera differing by 180° in azimuth. After the photographic plates are developed (exposure time 1 s), the star tracks and the intersection of the lines connecting the fiducial marks are measured on a comparator. The transformation of the plate coordinates into the α, δ-system provides the rectascension and the declination of the zenith point, after averages are taken and corrections are made for the reading of the levels. The observations (several plates) require one half to one hour, including instrumental setup and disassembly. The attainable accuracy is ±0.5″ or better (BIRARDI 1976, SEEBER and TORGE 1985).

Recent developments in the CCD (charge-coupled device) technique allow creation of an electronic image of the stars. The principle of a CCD is based on the photoelectric effect produced on a semiconductor plate. The number of collected photoelectrons is proportional to the light received. A CCD matrix is composed of a number of linear arrays. CCD's with 2000 x 2000

pixels and more are available with pixel dimensions of 10 μm and less and a resolution of 0.1 pixels. Using an image processing unit and a PC, an automatic evaluation of star observations is achieved, reaching similar accuracies as with photographic methods (KOVALEVSKY 1995, GERSTBACH 1996).

For *time determination*, an accuracy of ±1 ms is needed in optical geodetic astronomy. This is provided by quartz clocks, which are based on quartz crystal oscillators (frequency stability $10^{-8}...10^{-9}$ over a few hours) and are synchronized by time signals, cf. [2.2.2]. A simple time measurement is possible with a GPS receiver, cf. [5.2.5]. In order to record the time of a star transit through the horizontal or vertical thread of a telescope, or the time of taking a star image, a registration device (chronograph) has to be implemented in the observation system.

5.3.2 Astronomic Positioning and Azimuth Determination

The determination of astronomic latitude, longitude and azimuth is based on the relations given in [2.6.2], where the star positions (α, δ) are given by star catalogues, cf. [2.4.3]. We mention here only a few of the many methods developed in geodetic astronomy.

In determining the *astronomic latitude* Φ, it is required, according to (2.21), to ascertain the zenith angle z and the hour angle h. The latter quantity has to be derived from the rectascension α and the measured time, see (2.22). By differentiating (2.21) with respect to the observed quantities, we obtain differential formulas for estimating the effect of observational errors on the results and for finding optimum configurations for the observations. For the latitude we obtain

$$d\Phi = -\frac{1}{\cos A} dz - \cos \Phi \tan A \, dh . \tag{5.50}$$

An error in z has a minimum effect for meridian transits of the star, and an error in h has no effect on Φ in that case.

For an upper culmination (the smaller zenith angle) of a northern star ($A = 0°$) or a southern star ($A = 180°$), the latitude is given by (see also Fig. 2.4)

$$\Phi = \delta_N - z_N \text{ and } \Phi = \delta_S + z_S \tag{5.51}$$

respectively. Measuring the meridian zenith angle (e.g., to Polaris) is therefore most suitable for the determination of the latitude.

If one observes a pair of stars consisting of both a northern and a southern star having approximately the same zenith angle, averaging (5.51) significantly reduces the uncertainties in z due to refraction (*Sterneck* method):

$$\Phi = \frac{1}{2}(\delta_N + \delta_S) + \frac{1}{2}(z_S - z_N).$$ (5.52)

In the *Horrebow-Talcott* method, the small difference between the meridian zenith angle of the northern and southern stars in a star pair is measured by a registering micrometer. The optical axis in each case is adjusted to the same zenith angle using the Horrebow level that is mounted on the horizontal axis. Since accurate circle and time readings are not required, this method provides very precise results ($\pm 0.1''$ when about 20 star pairs are observed).

The *astronomic longitude* Λ is given by the difference between the local sidereal time LAST and the Greenwich sidereal time GAST (2.23):

$$\Lambda = LAST - GAST,$$ (5.53)

where 1 s corresponds to $15''$. According to (2.22), LAST is related to the hour angle h:

$$LAST = h + \alpha.$$ (5.54)

h can be computed from the zenith angle according to (2.21). If the latitude is known, we have:

$$\cos h = \frac{\cos z - \sin \Phi \sin \delta}{\cos \Phi \cos \delta}.$$ (5.55)

Converting measured universal time UT to GAST allows the determination of Λ.

By differentiating (2.21), we get

$$dh = -\frac{dz}{\sin A \cos \Phi} - \frac{\cot A}{\cos \Phi} d\Phi.$$ (5.56)

The effect due to errors in z is minimum when observations are made on the prime vertical ($A = 90°$), while the effect here is zero for errors in Φ. The influence of refraction is largely eliminated when observing east and west stars of the same altitude that are symmetric with respect to the meridian. On the other hand, by observing the time of transit across the meridian ($h = 0$) we get

LAST = α. An accuracy of ± 0.01 to 0.02 s is obtained from approximately 30 transits.

The accuracy of the determination of longitude depends primarily on the systematic errors of the observer, the instrument, and the time comparison. If the determinations are made by the same observer, using the same instrument and the same time signal transmitting station, as well as the same stars, then longitude differences are essentially free from these errors. Longitude determinations of high accuracy have thus been carried out as measurements of differences with respect to a reference station.

An economical method to determine the latitude and longitude simultaneously is known as the method of *position lines*.
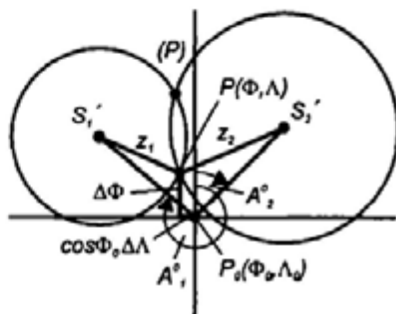

Fig. 5.27. Method of position lines

The zenith angles z_1, z_2 of two stars $S_1(\alpha_1, \delta_1)$, $S_2(\alpha_2, \delta_2)$ are observed at sidereal times LAST_1, LAST_2 and at azimuths A_1, A_2. If S_1, S_2 are projected on the earth's surface, then the intersections of the circles centered at the projections S_1', S_2' having radii z_1, z_2 represent two geometric positions P and (P) for the point of observation (Fig. 5.27). Near P, the circles can be replaced by their tangent lines (position lines). The intersection of these lines yields an approximation to P. Computationally, one obtains the corrections $\Delta\Phi = \Phi - \Phi_0$ and $\Delta\Lambda = \Lambda - \Lambda_0$ upon introducing an approximate position $P_0(\Phi_0, \Lambda_0)$. When observations are made with the prism astrolabe, cf. [5.3.1], the zenith angle predetermined by the prism is treated as an additional unknown. One obtains accuracies of a few $0.1''$ from about 20 stars evenly distributed above the horizon.

A *zenith camera* also permits the simultaneous determination of latitude and longitude. From the coordinates α_z, δ_z of the zenith, derived from the observations, we obtain

$$\Phi = \delta_z, \quad \Lambda = \alpha_z - \text{GAST}, \tag{5.57}$$

cf. [5.3.1], see also Figs. 2.2, 2.4.

If the latitude is known, then the azimuth A can be obtained, according to (2.21), from the hour angle h derived from the sidereal time and the rectascension (5.54):

$$\tan A = \frac{\sin h}{\sin\Phi\cos h - \cos\Phi\tan\delta}. \tag{5.58}$$

Differentiation of (2.21) yields

$$dA = \frac{\cos q\cos\delta}{\sin z}dh + \cot z\sin A\,d\Phi, \tag{5.59}$$

where q is the parallactic angle, cf. [2.6.2]. An error in h has a minimum effect for $\delta \approx 90°$ (stars near the pole). For some 10 observations, we get an accuracy of $\pm0.3''$ to $0.5''$. The azimuth of a terrestrial target is obtained by additionally measuring the angle between the directions to the star and the target.

5.3.3 Reductions

In order to refer the "observed" positions (epoch t) of fixed stars to the system of the star catalogue (mean positions at the reference epoch t_0), several reductions have to be applied.

* *Astronomic refraction* causes an apparent increase in the star's altitude (Fig. 5.28). The true zenith angle z is obtained from the observed quantity z' by adding the astronomic refraction angle $\Delta z_{_{\cdots}}$:

$$z = z' + \Delta z_{_{\cdots}}. \tag{5.60}$$

According to (5.10) and (5.11), the refraction angle depends on the vertical gradient of the refractive index along the path. For a standard atmosphere (temperature 288.15 K, atmospheric pressure 1013.25 hPa), we obtain for $z' < 70°$:

$$\Delta z_0 = 57.08''\,\tan z' - 0.067''\,\tan^3 z'. \tag{5.61}$$

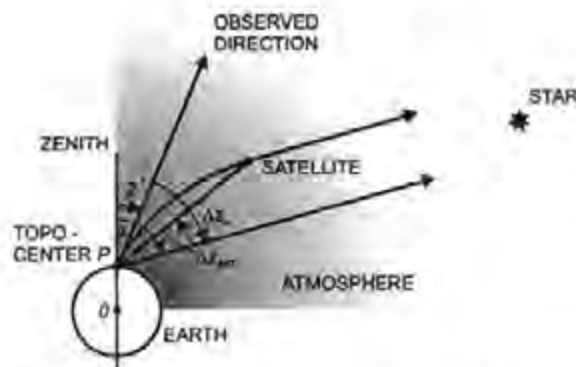For actual conditions (temperature T, pressure p), we have the transformation

Fig. 5.28. Astronomic and satellite refraction

$$\Delta z_{\pi} = \Delta z_0 \frac{p}{1013} \frac{288}{T},$$
(5.62)

(SAASTOMAINEN 1972/1973, KOVALEVSKY 1995). The accuracy of Δz_{π} varies between a few $0.01''$ and a few $0.1''$ and depends strongly on systematic deviations from the atmospheric model (turbulences, slope of the atmospheric layers).

- The *diurnal aberration* is an apparent displacement in direction resulting from the finite velocity of light and the relative velocity of the observer with respect to the stars, due to the earth's rotation. The corresponding reduction reaches a maximum of $0.3''$ at the equator.
- The *geocentric* (or diurnal) *parallax* represents the difference between the topocentric and the geocentric direction; it can be neglected for star observations, cf. [2.4.1].

Through these reductions, the "observed" position is transformed to the "apparent" position (apparent place) at epoch t. The reduction from the "mean" position (epoch t_0) to the apparent position (epoch t) involves the following steps:

- Applying *precession* and *proper motion* for the time interval $t - t_0$ transforms the mean position (t_0) to the mean position at epoch t, cf. [2.4.2], [2.4.3].
- Accounting for *nutation* transforms the mean position (t) to the true position (t). The origin of the system is still at the barycenter of the solar system.
- The transition to the (geocentric) apparent position (t) is performed by accounting for the apparent directional changes arising from the orbital motion of the earth around the sun (*annual aberration*, up to $20''$) and the

difference between the heliocentric and the geocentric directions (*annual parallax*, less than 1").

Instead of reducing from the mean position (t_0) to the apparent position (t), astronomic almanacs that contain the apparent places of fundamental stars for a particular year (Apparent Places of Fundamental Stars, Astronomisches Recheninstitut Heidelberg) can be used.

Finally, we must consider that the results (astronomic positions and azimuths) refer to the instantaneous spin-axis of the earth. The results must be transformed into the IERS reference pole (ICRF) by applying reductions for polar motion, cf. [2.5.2]. Multiplying the polar motion rotation matrix (2.18) with the unit vector of the local vertical (2.19) gives the reductions for latitude, longitude, and azimuth (MUELLER 1969, p.86):

$$\Delta\Phi_P = \Phi_{ICRF} - \Phi = -\left(x_p \cos\Lambda - y_p \sin\Lambda\right)$$

$$\Delta\Lambda_P = \Lambda_{ICRF} - \Lambda = -\left(x_p \sin\Lambda + y_p \cos\Lambda\right)\tan\Phi, \tag{5.63}$$

$$\Delta A_P = A_{ICRF} - A = -\left(x_p \sin\Lambda + y_p \cos\Lambda\right)\sec\Phi$$

where x_p, y_p are the pole coordinates with respect to ICRF.

5.3.4 Very Long Baseline Interferometry

Extragalactic radio sources (quasars = quasi-stellar radio sources, radio galaxies) emit waves in the cm to dm range, which can be detected by large antennas (radio telescopes) used in radio astronomy. The approximate angular resolution of such a telescope is given by the wavelength/diameter ratio, and thus it is limited to a few arcmin for telescope diameters less than 100 m. By employing a receiving system of two widely (a few 1000 to 10 000 km) separated radio telescopes (baseline), the resolution can be increased to ±0.001" and better: *Very Long Baseline Interferometry* (VLBI), MORITZ and MUELLER (1987, p.381).

The wave train from an extragalactic radio source arrives at the telescope P_2 with a phase difference Φ with respect to the telescope P_1. Φ is related to the time delay τ, the time the wave requires to travel the path difference $c\tau$ (c = velocity of light in vacuum), Fig. 5.29. Due to the earth's rotation, Φ and τ depend on time t. The following relations are valid:

$$\Phi(t) = 2\pi\frac{c}{\lambda}\tau(t) = 2\pi\nu\tau(t), \tag{5.64}$$
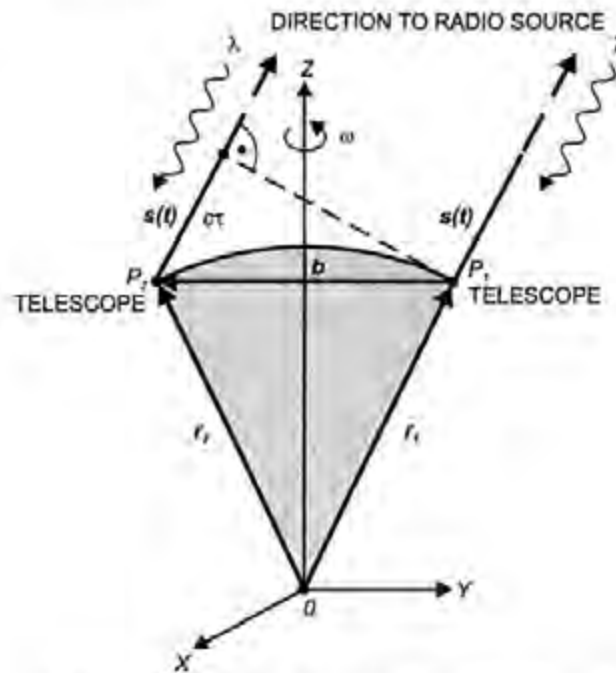
Fig. 5.29. Very Long Baseline Interferometry

where λ and ν are the respective wavelength and the frequency of the received radio wave. We introduce the baseline vector

$$\mathbf{b}_{ITRS} = \mathbf{r}_2 - \mathbf{r}_1 = \begin{bmatrix} X_2 - X_1 \\ Y_2 - Y_1 \\ Z_2 - Z_1 \end{bmatrix}, \tag{5.65}$$

described in the terrestrial geocentric system (2.14) and the unit vector to the quasar:

$$\mathbf{s}_{ICRS} = \begin{bmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{bmatrix}, \tag{5.66}$$

given in the celestial reference system (2.11). After transformation of **s** to the terrestrial system according to (2.16), we obtain, see Fig. 5.29,

$$\tau(t) = -\frac{1}{c}\mathbf{b}\cdot\mathbf{s}(t). \tag{5.67}$$

The negative sign takes the direction of **s** into account, which is opposite to the direction of wave propagation. By comparing the two wave trains received at P_1 and P_2, interferences are obtained. The *frequency* of the interference *fringes* (maxima and minima) changes due to the earth's rotation:

$$f(t) = \frac{1}{2\pi}\frac{d\Phi(t)}{dt}.$$ (5.68)

With (5.64) and (5.67), the fringe frequency can be expressed as

$$f(t) = v\frac{d\tau(t)}{dt} = -\frac{v}{c}\mathbf{b}\cdot\dot{\mathbf{s}}(t),$$ (5.69)

$\dot{\mathbf{s}} = d\mathbf{s}/dt$. Equations (5.67) and (5.69) represent the VLBI *observation equations* (CAMPBELL and WITTE 1978, CARTER et al. 1985).

The VLBI observables are the time delay τ (or the phase delay, respectively) and the delay rate $d\tau/dt$. They are derived by a comparison of the signals, which are recorded on magnetic tape, along with the precise time provided by hydrogen maser frequency standards. Typically, 10 to 20 sources are tracked over periods of 3 to 6 minutes several times over a 24-hour observation session. The data tapes are sent to a correlator center where they are replayed and the signals processed (MARK III and recently MARK IV correlator, CLARK et al. 1987). The correlation function furnishes the delay and the delay rate between the two stations. The observable τ represents a group delay, while Φ is a phase delay, cf. [5.1.1], which involves the problem of ambiguity resolution, cf. [5.2.5]. The fringe frequency (delay rate) observation is free of this problem but of less importance due to its lower accuracy as compared to the delay observation and since it allows only the determination of a reduced set of parameters (SEEBER 1993, p.430).

Reductions are applied for the daily aberration, cf. [5.3.3], for systematic clock differences (clock synchronization), for the effects of the tropospheric refraction, cf. [5.1.2], and for relativistic effects. The effect of the ionosphere is compensated for by observing in two frequency-bands, namely 2.2 and 8.4 GHz, cf. [5.1.3]. Main *error* sources result from timing (± 1 ps) and frequency instabilities ($\pm 10^{-15}$ over a few days) as well as from tropospheric models. The determination of the "wet" component plays a major part in this aspect, cf. [5.1.2]. Attempts have been made to measure the water vapor content along the signal path by water vapor radiometers in order to estimate the wet component with cm-accuracy. The vertical baseline component is mainly affected by the uncertainty of the wet component.

Among the *parameters* to be estimated from (5.67) and (5.69) are the components of the baseline vector in the terrestrial reference system (cm accuracy). Global solutions for a given epoch deliver a precision of a few mm for the coordinates and annual station velocities accurate to a mm/a. They also yield earth rotation parameters, cf. [2.5.3]. These parameters enter into the models through the transformation between the terrestrial and the celestial system. They can be determined with an accuracy better than 0.001″ (pole coordinates) and 0.1 ms (UT1), with a one-day resolution, cf. [2.5.2]. In addition, corrections to the precession and nutation models can be derived from VLBI (ROBERTSON et al. 1985, CAMPBELL et al. 1992).

Fig. 5.30. 20m-radio telescope, Fundamentalstation Wettzell, courtesy Bundesamt für Kartographie und Geodäsie (BKG), Frankfurt a.M., Germany

Today, around 20 stationary radio telescopes participate in international programs in order to establish the reference frames, to determine the earth rotation parameters, and to derive station velocities, especially within the International Earth Rotation Service. The stations are mainly located in North America, Europe, and Japan, Fig. 5.30. Mobile radio telescopes have been developed for rapid surveying of regions with recent crustal movements, such as California or the Eastern Mediterranean , and to fill gaps of the terrestrial reference system in the southern hemisphere.

5.4 Gravimetry

Gravimetry deals with the measurement of the gravity intensity (gravity) and the gravity gradient by terrestrial methods on or close to the earth's surface (MARSON and FALLER 1986, TORGE 1989). "Absolute" gravity measurements refer directly to the standards of length and time [5.4.1], while "relative" measurements use a counterforce for the determination of gravity differences [5.4.2]. A global gravity reference system is needed in order to refer local and regional gravity networks to a common standard [5.4.3]. Gravity measurements on moving platforms are valuable for areas difficult to access [5.4.4]. Local gravity-field information can be obtained by the measurement of the gravity gradient [5.4.5]. The continuous record of gravity provides gravity variations with time, especially the gravimetric earth tides [5.4.6].

The *unit* of gravity in the SI-system is ms^{-2}. The units $mGal = 10^{-5} ms^{-2}$ and $\mu Gal = 10^{-8} ms^{-2} = 10 nms^{-2}$ are still in widespread use in geodesy and geophysics and derived from the unit Gal (after Galileo) of the former cgs-system.

5.4.1 Absolute Gravity Measurements

An "absolute" gravity measurement determines the gravity g from the fundamental acceleration quantities length and time. We distinguish between the pendulum and the free-fall method, both introduced by Galileo Galilei (1564 – 1642), FALLER and MARSON (1988), TORGE (1991).

The pendulum method is no longer applied today but governed gravimetry for about 300 years. Because of its fundamental importance, and because recent results are still part of some gravity networks, a short introduction is given here.

The *pendulum method* is based on the measurement of the period and the length of a freely swinging pendulum. For a *mathematical* pendulum (point mass m suspended on a weightless wire of length l) we have the equation of oscillation

$$ml\ddot{\varphi} + mg\sin\varphi = 0, \tag{5.70}$$

with the phase angle $\varphi = \varphi(t)$, and the angular acceleration $\ddot{\varphi} = d^2\varphi/dt^2$ (Fig. 5.31). Integration over a full period leads to an elliptical integral. After expansion into a series, we obtain the period T of oscillation

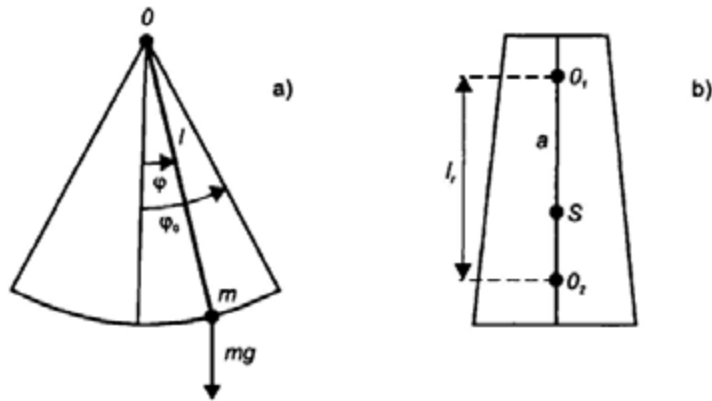$$T = 2\pi\sqrt{\frac{l}{g}}\left(1 + \frac{\varphi_0^2}{16} + \dots\right), \tag{5.71}$$

Fig. 5.31. Absolute pendulum method: a) mathematical pendulum, b) reversible pendulum

where the amplitude φ_0 is kept small. Thus, gravity is derived from the measurement of T and l.

A mathematical pendulum is difficult to realize. Equations (5.70) and (5.71) also hold for a *physical* pendulum if l is replaced by the reduced pendulum length

$$l_r = \frac{J}{ma}.$$

(5.72)

Here, J is the moment of inertia with respect to the axis of rotation O, m the total mass, and a the distance between O and the center of mass. The *reversible pendulum* is characterized by two axes of rotation, for which, after a corresponding adjustment, the same oscillation time is achieved. The distance between the two axes is equal to the reduced pendulum length, thus avoiding the direct determination of J, m, and a (Fig. 5.31).

The reversible pendulum was introduced by *Kater* (1818), and a limited number of observations were carried out, primarily after the 1860's (transportable devices by *Repsold* and others). After the fundamental gravity determination in Potsdam, cf. [5.4.3], only a few more experiments were performed. Due to errors in determining the length of the swinging pendulum, the accuracy achieved remained at a few μms^{-2} (SCHÜLER et al. 1971).

The *free-fall method* is based on the equation of motion

$$m\ddot{z} = mg(z)$$

(5.73)

of a freely falling body. Here m is mass; z is along the local vertical axis, and $\ddot{z} = d^2z/dt^2$ (Fig. 5.32). Assuming a homogeneous gravity field along the falling distance, double integration of (5.73) yields the free-fall equation

$$z = z_0 + \dot{z}_0 t + \frac{g}{2}t^2 .\tag{5.74}$$

Equation (5.74) relates the position z of the falling body at the time t to gravity. The integration constants z_0 and $\dot{z}_0$ represent z and $\dot{z} = dz/dt$ at the starting time of the experiment ($t = 0$).
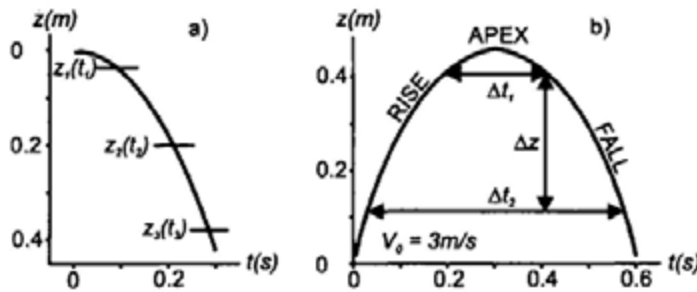

Fig. 5.32. Distance-time diagram: a) free-fall method, b) symmetrical rise and fall

z_0 and $\dot{z}_0$ slightly deviate from zero due to problems in accurately defining the starting position and small microseismic accelerations. The gravity change along the falling distance (non-homogeneous gravity field) is taken into account by introducing the vertical gradient $\partial g/\partial z$ into (5.74) and by referring the final gravity value to a fixed reference height, e.g. 1 m (NIEBAUER 1989).

If at least three position/time pairs are measured, z_0 and $\dot{z}_0$ can be eliminated in (5.74), and gravity is given by

$$g = 2\frac{(z_3 - z_1)(t_2 - t_1) - (z_2 - z_1)(t_3 - t_1)}{(t_3 - t_1)(t_2 - t_1)(t_3 - t_2)} .\tag{5.75}$$

For the *symmetrical rise and fall*, the test mass is thrown vertically upward and falls back after having reached the apex (Fig. 5.32). It is sufficient to measure time at the same two positions during rise and fall. Evaluation of (5.74) yields

$$g = \frac{8\Delta z}{\Delta t_2^2 - \Delta t_1^2} ,$$

$$(5.76)$$

with $\Delta z = z_2 - z_1$, $\Delta t_1 = t_3 - t_2$, $\Delta t_2 = t_4 - t_1$.

Generally, more than the necessary number of position/time pairs is measured. Equation (5.74) then serves as observation equation for a least squares adjustment of g, z_0, and $\dot{z}_0$.

Accuracy demands for absolute gravimetry are at the order of 10^{-9} g or 10^{-8} ms^{-2}. Hence, for a falling distance of 0.2 m (falling time 0.2 s), accuracies of ±0.2 nm and ±0.1 ns are required for position and time respectively. This is achieved by interferometric distance measurements and simultaneous electronic timing.
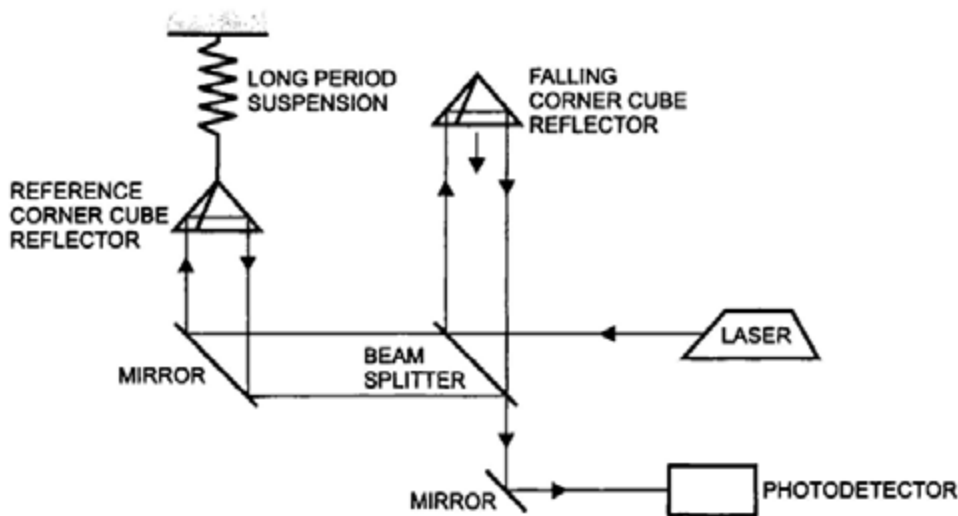

Fig. 5.33. Michelson interferometer principle

For recent free-fall gravimeters, a polarization or iodine stabilized He-Ne gas laser ($\lambda = 633$ nm) serves as the length standard and an atomic (rubidium) frequency normal as the time standard. A Michelson interferometer is used for the distance measurement, with two corner cube reflectors as the main components (Fig. 5.33). One of the reflectors is fixed and serves as a reference; the other one represents the falling body. By splitting the laser light into a measurement and a reference beam, and superimposing them again after parallel reflection, light interferences occur. The zero crossings of this fringe signal have a distance of $\lambda/2$, and the fringe frequency increases with time due to the velocity increase according to $\dot{z}(t) = gt$ (Fig. 5.34). The zero crossings are

sensed by a photodiode, converted to an electronic signal, amplified, triggered, and counted. A time measurement (atomic clock and time interval counter) is carried out after a preset number n of zero crossings, which corresponds to a falling distance of

$$\Delta z = n\frac{\lambda}{2}. \tag{5.77}$$

The experiments are performed in vacuum (10^{-4} Pa) in order to eliminate air resistance. Microseismicity is to a large part absorbed by long-period ($T > 10$ s) compensation devices. A further reduction is achieved by randomization, performing a large number (several 100 to a few 1000) of drops per station. For the rise and fall method, systematic errors that are proportional to the falling body's velocity (residual air drag, timing errors) cancel.
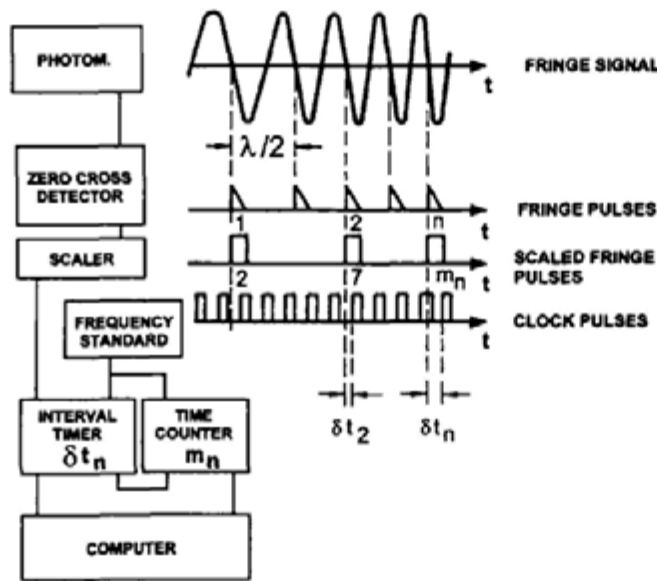

Fig. 5.34. Timing of scaled fringe pulses, after ZUMBERGE (1981)

Several *reductions* have to be applied to the observed gravity values. The gravimetric earth tides can be reduced with an accuracy of a few $0.01\,\mu ms^{-2}$ in most parts of the world, cf. [8.3.5]. The *polar motion* reduction, according to (3.109) and (5.63), is given by

$$\delta g_{\text{pole}} = -\delta_{\text{pole}}\omega^2 R\sin 2\varphi\left(x_p\cos\lambda - y_p\sin\lambda\right), \tag{5.78a}$$

with ω = rotational velocity of the earth, R = earth radius, and x_p and y_p = coordinates of the instantaneous pole with respect to the IERS reference pole.

The geodetic coordinates φ,λ sufficiently approximate astronomic latitude and longitude. The factor $\delta_{pole} = 1.2$ accounts for the earth's elasticity, cf. [5.4.6]. The direct (gravitation) and indirect (deformation) effect of *air pressure* variations is taken into account by a reduction

$$\delta g_p = 3(p - p_n)\text{nms}^{-2},\qquad(5.78b)$$

with p = actual air pressure, and p_n = normal air pressure as given by a standard atmosphere, both in hPa (NIEBAUER 1988). The reduction of the adjusted gravity value from the reference height to ground is performed by relative gravity measurements, with an accuracy of ±0.01 to 0.02 μms^{-2}, cf. [5.4.5].

The long-term *stability* of the length and time standards is controlled by calibration of the laser (10^{-9} to 10^{-10} frequency stability) and the atomic clock (10^{-10}). The *repeatability* of the gravimeter system can be checked by regular measurements at a reference station (Fig. 5.35), while the *accuracy* should be estimated by comparisons with other instruments, cf. [5.4.3].
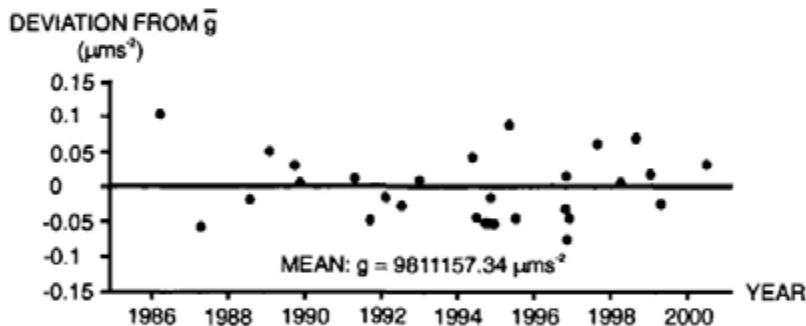

Fig. 5.35. Long-term stability control of JILAG-3 absolute gravimeter at the reference station Clausthal, Germany

The *accuracy* of absolute gravity measurements depends strongly on site conditions. Stable sites (hard bedrock, low man-made noise) provide better results than locations in sediments, close to the coast, or in urban environment. The drop-to-drop scatter (±0.05 to ± a few μms^{-2}) is reduced by a large number of measurements. The adjusted station gravity-value is generally derived with a standard deviation of ±0.01 to 0.03 μms^{-2}. The accuracy is at the order of a few 0.01 μms^{-2}, due to unmodeled systematic effects of the instrument (e.g., floor recoil) or the environment (atmospheric loading, ground water variations). Systematic discrepancies between different instruments may reach 0.05 μms^{-2} and more, cf. [5.4.3].

The free-fall method was developed in the 1950's (*Volet* and *Sakuma* at the BIPM Sèvres, *Cook* at the National Physical Laboratory, Teddington). A transportable instrument was designed and employed by *Faller* in 1968. Among the presently operating transportable gravimeters are the JILA (Joint Institute for Laboratory Astrophysics, Boulder, CO, U.S.A.) and the FG5 (Micro-g Solutions Inc., U.S.A.) free-fall instruments (FALLER et al. 1983, NIEBAUER et al. 1995). Around 200 position/time data pairs are collected over one drop, evenly distributed in distance over the drop length of 20 cm, and adjusted on-line to a fitting parabola. The falling object moves in a co-accelerated 'drag-free' chamber. The chamber eliminates residual air drag and serves, by adequate acceleration, for dropping and catching the corner cube as well as for transporting it back to the initial position. The reference corner cube is isolated from ground motions by a "super-spring", which by a feed-back system electronically generates effective free-oscillation periods between 30 and 60 s. While the JILA gravimeters have a horizontal interferometer basis, the FG5 instruments employ a vertical basis (Figs. 5.36, 5.37). The vertical basis eliminates the influence of floor vibration and tilt on the optical path length. The instruments are disassembled for transportation (FG5: 240 kg in 8 containers). Setting up at a station requires about two hours, and observations are generally carried out over one to two days, depending on local noise (TORGE et al. 1987, KLOPPING et al. 1997). A portable modification of the FG5 gravimeter can be used in outdoor environment on quiet sites, providing a precision of $\pm 0.1 \, \mu ms^{-2}$ in 10 minutes.
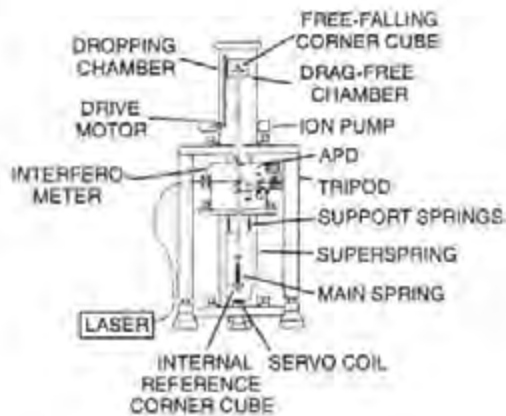
5.36. Free-fall gravimeter FG5 principle, courtesy Micro-g Solutions, Inc., Arvada, CO, U.S.A.

5.37. Free-fall gravimeter FG5 view, courtesy Micro-g Solutions, Inc., Arvada, CO, U.S.A.

Transportable *rise-and-fall* instruments have been developed by the Istituto di Metrologia "G.Colonnetti", Torino, Italy (ALASIA et al. 1982) and by Jaeger S.A., France (SAKUMA 1983). An option for the rise-and-fall mode is also available for the FG5 gravimeter.

An *ocean-bottom* absolute gravimeter has been constructed primarily for monitoring long-term gravity variations in geodynamically active zones (ZUMBERGE and CANUTESON 1995).

5.4.2 Relative Gravity Measurements

A "relative" gravity measurement yields the gravity difference between two stations or variations of gravity with time, cf. [5.4.6]. Either time or length is measured, keeping the other quantity fixed. As a consequence, relative measurements can be performed more easily than absolute ones.

For the *pendulum method*, the oscillation periods T_1 and T_2 of the same pendulum are measured at two stations P_1 and P_2. From (5.71) we obtain

$$\frac{g_2}{g_1} = \frac{T_1^2}{T_2^2} \tag{5.79}$$

or, after simple transformation, the gravity difference

$$\Delta g_{1,2} = g_2 - g_1 = -2g_1 \frac{T_2 - T_1}{T_2} + g_1 \frac{(T_2 - T_1)^2}{T_2^2}. \tag{5.80}$$

The relative pendulum method has been employed extensively since *v. Sterneck* (1887) developed a transportable device (pendulum length 25 cm, two pendulums swinging on the same support in opposite phase in order to eliminate floor recoil effects). Although the systematic effects that are independent of time and position cancel with this differential method, the accuracy could not be increased over a few $\mu m s^{-2}$ due to problems in keeping the pendulum length constant during a field survey. The method was superseded in the 1930's by elastic spring gravimeters. The pendulum method was used until the 1960's for establishing gravimeter calibration lines, exploiting the fact that pendulum results are given in the unit of acceleration and do not need to be calibrated.

Relative *gravity meters* use a counterforce in order to keep a test mass in equilibrium with gravity. Gravity changes in space or time are monitored by corresponding changes of the counterforce, which are transformed to the gravity unit by a calibration function. An elastic counterforce is used nearly exclusively, but magnetic counterforces are also employed in instruments operating on moving platforms and in the stationary mode, cf. [5.4.4], [5.4.6].

The elastic spring *gravimeter* is based on the principle of a spring balance. If gravity changes, the spring length will also change in order to maintain static equilibrium between gravity and the elastic force. According to *Hooke*'s law, the strain is proportional to the stress for small elongations. We distinguish between translational systems (seldom used) and rotational systems.

In a *translational system* (vertical spring balance), the condition of equilibrium is given by (Fig. 5.38a)

$$mg - k(l - l_0) = 0, \tag{5.81}$$

where k is the spring constant and l (resp. l_0) is the length of the spring with (resp. without) a load. Applying (5.81) on a gravity difference Δg furnishes a linear relationship between Δg and the observed difference in length Δl:

$$\Delta g = \frac{k}{m} \Delta l = \frac{g}{l - l_0} \Delta l. \tag{5.82}$$

An undamped spring generates a harmonic oscillation with the proper frequency

$$\omega_0 = \sqrt{k/m} \tag{5.83}$$

and the oscillation time

$$T_0 = 2\pi \sqrt{\frac{m}{k}} = 2\pi \sqrt{\frac{l - l_0}{g}}. \tag{5.84}$$

By differentiation, we obtain the mechanical *sensitivity*

$$\frac{dl}{dg} = \frac{m}{k} = \frac{T_0^2}{4\pi^2}. \tag{5.85}$$

In order to assess gravity changes with a relative accuracy of 10^{-8}, length changes of a 0.1 m long spring would have to be determined to ± 1 nm.

Rotational systems (lever spring balance) consist of a lever that supports a mass m and rotates about an axis O. Equilibrium can be produced through a horizontal torsion spring or through a vertically or obliquely acting helical spring. The equilibrium of the torques for the *lever torsion* spring *balance* (Fig. 5.38b) yields

$$mg\cos\alpha - \tau(\alpha_0 + \alpha) = 0, \tag{5.86}$$

where a = length of the lever, α = angle between the horizontal and the lever, τ = torsion constant, and α_0 = pretension angle of the spring. This non-linear system becomes a linear one for $\alpha = 0$, with

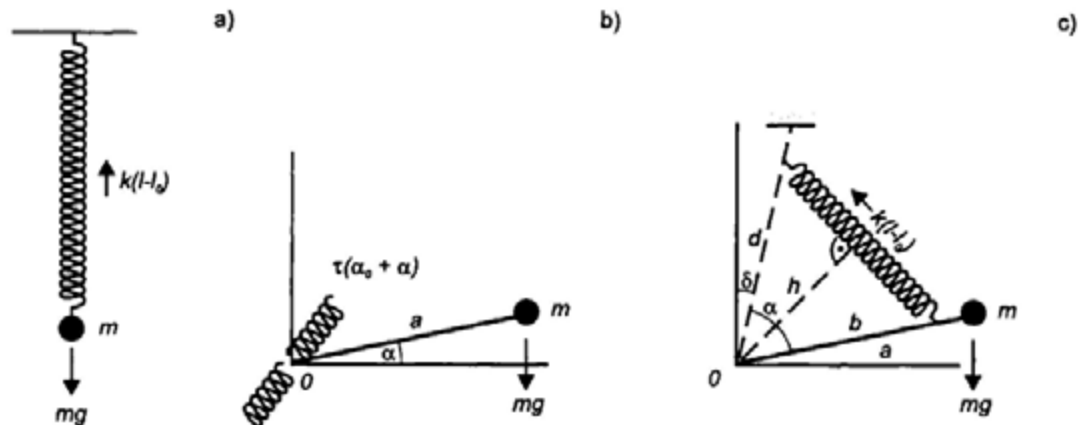$$\Delta g = \frac{\tau}{ma} \Delta \alpha. \tag{5.87}$$

Fig. 5.38. Elastic spring gravimeter principle: a) vertical spring balance, b) lever torsion spring balance, c) general lever spring balance

For the *general lever spring balance*, the spring counterforce acts under an arbitrary angle on the lever carrying the mass. The line connecting the rotation axis O with the point where the spring is mounted deviates by an angle δ from the vertical (Fig. 5.38c). With the distance

$$h = (bd/l)\sin\alpha \qquad (5.88a),$$

the equilibrium condition for the torques reads

$$mga\sin(\alpha+\delta) - kbd\frac{l-l_0}{l}\sin\alpha = 0. \qquad (5.88b)$$

The sensitivity of this non-linear system can be significantly increased by approximating the torques of gravity and of the elastic spring (astatization). With a zero-length spring ($l_0 = 0$), we have the sensitivity

$$\frac{d\alpha}{dg} = \frac{\sin(\alpha+\delta)\sin\alpha}{g\sin\delta}. \qquad (5.89)$$

High sensitivity is achieved at a small angle δ and $\alpha \approx 90°$. For $a = 0.1$ m, $\alpha+\delta = 90°$, and $\delta = 100''$, displacements have to be measured with ± 2 μm in order to obtain a relative accuracy of 10^{-8}. Compared to the linear system, the sensitivity is thus increased by a factor of 2000.

The required accuracies of 0.1 μms^{-2}, or better, place high demands on the reading systems as well as on the stability of the counterforce with time.

Optical and/or electrical reading systems are used to observe the position of the test mass. A capacitive position-indicator is usually employed and is connected to a digital readout unit. The *zero-method* is preferred for the measurement of the displacement, with a compensation device for restoring the zero position. Mechanical compensation is performed by a measurement screw, while electronic feedback systems are not affected by screw errors (RÖDER et al. 1988).

The *elasticity* of the spring should exhibit a time stability of 10^{-8} over several hours, which is the time interval required for transporting the gravimeter between the stations of a large-scale network, cf. [7.4]. Spring materials include NiFe alloys (small thermoelastic coefficient) and fused quartz (large but linear thermoelastic coefficient, small coefficient of thermal expansion). In addition, the measurement system has to be protected against changes in temperature (thermostat), air pressure (air-tight sealing), and magnetic field (shielding of metal alloy springs). The effects of mechanical shocks and vibrations can be reduced by a damping device, in addition to air-damping.

Spring gravimeters have been developed since the 1930's for use in geophysical exploration. From the 1950's, instruments were available which could be used for establishing large-scale gravity networks. Most of these early gravimeters had a limited measuring range (e.g. 2000 μms^{-2}) and required a reset screw for changing to another gravity range (Askania gravimeter: torsion spring balance, metal alloy; Worden gravimeter: fused quartz system with horizontal beam and vertical counter spring). The LaCoste and Romberg astatized gravimeters employ a metal zero-length spring, acting at 45° inclination on the horizontal beam (model G: 70 000 μms^{-2} range, measuring screw with 10 μms^{-2} per one rotation), KRIEG (1981), KANNGIESER (1983), Fig. 5.39. Recently developed instruments are microprocessor-controlled and are highly automated. They employ capacitive transducers and electronic feedback systems with worldwide
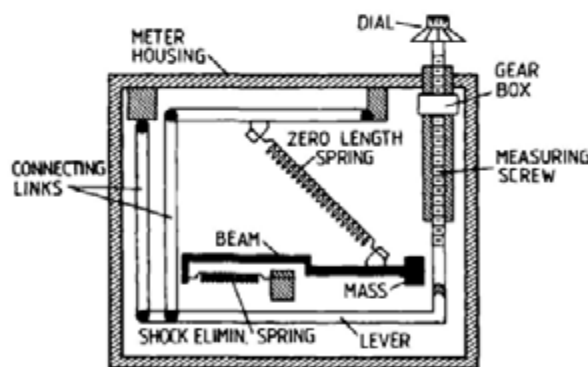

Fig. 5.39. LaCoste and Romberg gravimeter principle, courtesy LaCoste and Romberg, Inc., Austin, Texas, U.S.A.

Fig. 5.40. LaCoste and Romberg gravimeter (Graviton-EG), courtesy LaCoste and Romberg, Inc., Austin, TX, U.S.A

Fig. 5.41. Scintrex Autograph CG-3M gravimeter, courtesy Scintrex, Concord, Ontario, Canada

range. Self-leveling, a high data acquisition rate, and on-line evaluation (data compression and analysis, earth tides reduction, drift control) are further characteristics (Scintrex CG-3 Autograph: Worden type quartz system; LaCoste and Romberg model E-meter), FALK (1995), BONVALOT et al. (1998), Figs. 5.40, 5.41.

Options of conventional land gravimeters include underwater and bore-hole instruments. After sealing in a pressure and water protected diving bell, an *underwater gravimeter* is lowered to the sea bottom and remotely operated from on board a survey vessel. The underwater gravimeters are used mainly in the shelf areas, BEYER et al. (1966). Bore-hole gravimeters are characterized by small dimensions and remote-controlled operation at high temperatures.

Air/sea gravimeters will be descibed in [5.4.4] and recording (earth tides) gravimeters in [5.4.6].

Despite all measures to protect the gravimeter's measuring system against environmental disturbances, the zero reading changes with time: drift and tares. The *drift* is caused by aging of the spring material (approximating zero after some years) and short-term changes which occur during a field survey. This "transportation" drift results from reactions of the spring to vibrations and small shocks, uncompensated temperature fluctuations, and elastic aftereffects after unclamping the lever. It depends on the spring material and on measurement conditions and can reach a few μms^{-2}/day. Larger mechanical shocks may produce sudden *tares* of the same order of magnitude or more. The drift is determined by repeated station occupations during one day and subsequent modeling. Different methods have been developed depending on the instrumental behavior and the network structure; among them are the profile, the star, and the step method (Fig. 5.42).

After reducing the gravimeter reading for the earth tides, cf. [8.3.5], the drift function can be modeled by a low-order polynomial with time (Fig. 5.43):
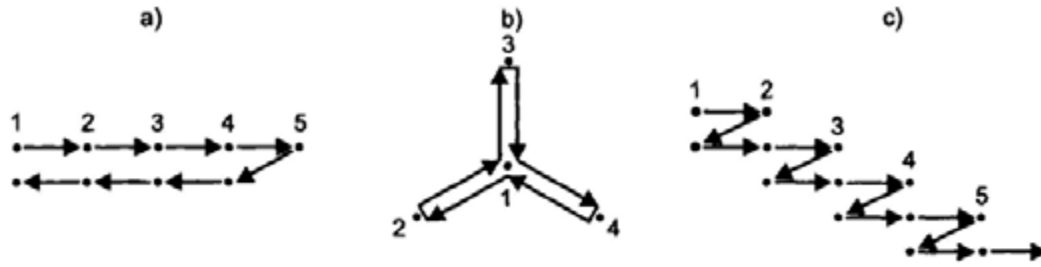
Fig. 5.42. Drift determination methods: a) profile method, b) star method, c) step method

$$D(t) = d_1 (t - t_0) + d_2 (t - t_0)^2 + \ldots, \qquad (5.90)$$

with t_0 = starting time (e.g., beginning of the survey) and d_1, d_2 = drift parameters. The network adjustment delivers the drift parameters, based on repeated observations, cf. [7.4].

The gravimeter reading z (in counter units) is converted to the gravity unit by means of the *calibration* function:

$$g = F(z). \qquad (5.91)$$

$F(z)$ depends on the physical and geometrical parameters of the measuring system, see (5.82), (5.87), (5.89), which cannot be determined with the desired
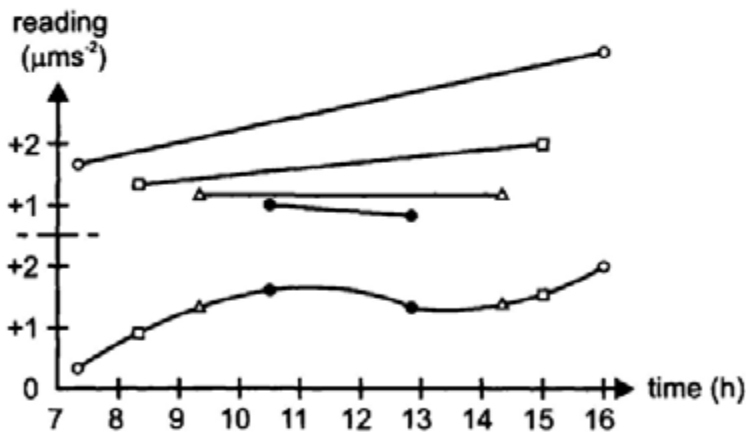

Fig. 5.43. Drift determination (profile method)

accuracy. Therefore, the calibration function is derived by comparing reading differences with known differences of gravity. A low degree polynomial (linear "scale factor" and eventually small non-linearities due to the properties of the feed-back system) provides the transformation from the readings (stations i, j) to the gravity difference:

$$\Delta g_{ij} = g_j - g_i = Y_1\left(z_j - z_i\right) + Y_2\left(z_j^2 - z_i^2\right) + \dots, \tag{5.92}$$

with Y_1, Y_2 = calibration coefficients. Periodic calibration terms may be added for taking cyclic errors of a measuring screw into account.

Laboratory and field methods are available for determining the coefficients of the calibration function. In the *laboratory*, gravity changes can be simulated and compared with the corresponding gravimeter readings. The tilt-table method uses the inclination by a known angle for producing an apparent gravity variation, and the mass method uses the defined change of the gravimeter mass. Special methods have been developed for recording gravimeters, cf. [5.4.6]. *Calibration lines* provide gravity differences, determined by absolute gravimeters, and may be densified by relative gravimetry. The calibration-line surveys exploit the fact that gravity varies with latitude and height (KANNGIESER et al. 1983). The limited gravity range of these lines only allows determination of an approximate value for the linear calibration factor; an improved estimate must be based on a global gravity-reference-system, cf. [5.4.3]

The *accuracy* of gravity differences ($\Delta g < 1000$ to 2000 μms^{-2}) observed with well calibrated and drift-controlled instruments is ± 0.1 to 0.2 μms^{-2}. Repeated measurements and the use of several instruments increase the accuracy to ± 0.05 to 0.1 μms^{-2} and ± 0.02 μms^{-2} for local ties (TORGE 1984, BECKER et al. 2000).

5.4.3 Gravity Reference Systems

Gravity reference systems provide homogeneity of gravimetric surveys by realizing a gravity standard through the gravity values of a selected number of stations.

The need to establish a global reference system arose at the end of the 19th century when larger sets of absolute and relative pendulum measurements had to be combined. *The Potsdam Gravity System* was introduced in 1909 by IAG. It was based on reversible pendulum measurements carried out in the Geodetic Institute Potsdam by *Kühnen* and *Furtwängler* (1898 – 1904). Relative pendulum ties to national base stations transferred the absolute value to other parts of the world. Since the 1930's, new absolute and relative gravity measurements revealed that the Potsdam gravity value was 140 μms^{-2} too high and that transfer errors of several 10 μms^{-2} had occurred.

The Potsdam Gravity System was superseded by the *International Gravity Standardization Net 1971* (IGSN71), recommended by the I.U.G.G. (MORELLI et al. 1974). This network contains 1854 gravity stations (among them about
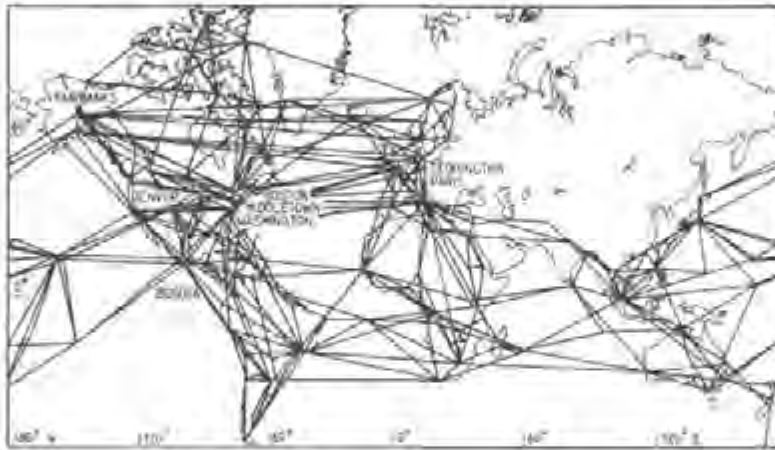
Fig. 5.44. International Gravity Standardization Net 1971 (IGSN71): absolute gravity stations and selected network ties, after MORELLI et al. (1974)

500 primary stations) determined by 10 absolute and about 25 000 relative measurements, including 1200 relative pendulum ties (Fig. 5.44). The mean uncertainty of the adjusted gravity values is less than $\pm 1\ \mu ms^{-2}$. High relative-accuracy is provided at gravimeter calibration lines, which extend in the north-south direction in America, Europe and Africa, and in the western Pacific. Meanwhile the IGSN71 has been extended to previously uncovered parts of the world. Regional networks have been connected to IGSN71, or transformed (shift and scale factor) to it, with the help of identical stations. The IGSN71 gravity values can be used to derive the linear calibration factor of relative gravimeters with a relative accuracy of some 10^{-5}.
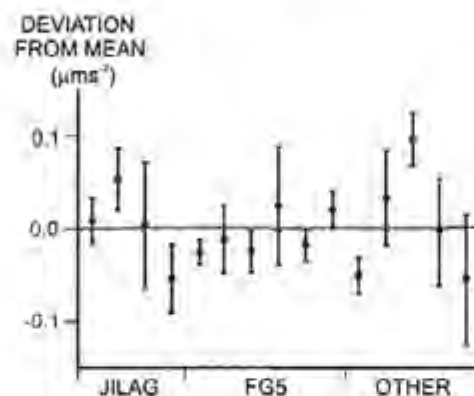

Fig. 5.45. Absolute gravimeter comparison, BIPM Sevres 1997: Deviations from mean value and standard deviations for JILAG-, FG5- and other gravimeters, after ROBERTSSON et al. (2001).

With the increasing availability of transportable *absolute gravimeters* of accuracy of 0.05 μms^{-2} and better, the gravity reference can be established independently from a global system with any gravity survey (TORGE 1998). The linear calibration factors of relative gravimeters then are derived from the absolute values available in the survey area. The IGSN71 consequently does not need a readjustment but rather is improved continuously by networks based on absolute gravimetry. This strategy requires a regular quality control of the absolute gravimeter systems, cf. [5.4.1]. International comparisons have been carried out since the 1980's at the BIPM, Sèvres (Fig. 5.45). For advanced absolute gravimeters, both the r.m.s. scatter around the BIPM mean value and the long-term stability is a few 0.01 μms^{-2}, which characterizes the present state of the realization of the gravity standard (ROBERTSSON et al. 2001).

5.4.4 Gravity Measurements on Moving Platforms

Kinematic methods have been developed for rapid and high-resolution gravimetric surveys in areas of having challenging environmental conditions, such as oceans, the polar regions, high mountains, and tropical forests. Ships and airplanes are predominantly used as platforms, but helicopters and land vehicles also have been employed for local surveys. Compared to stationary gravimetry, additional difficulties arise in kinematic gravimetry, including problems with orientation of the gravity sensor and with separating gravity from non-gravitational accelerations (DEHLINGER 1978, BROZENA and PETERS 1995, CANNON and LACHAPELLE 1997).

The principle of kinematic gravimetry is based on Newton's *equation of motion*. In the local astronomic system (also called local level system), cf. [2.6.2], the gravity vector is expressed by:

$$\mathbf{g}^l = \ddot{\mathbf{r}}^l - \mathbf{R}_b^l \mathbf{f}^b + \left(2\boldsymbol{\omega}_{ie}^l + \boldsymbol{\omega}_{el}^l\right) \times \dot{\mathbf{r}}^l , \tag{5.93}$$

where $\ddot{\mathbf{r}} = d^2\mathbf{r}/dt^2$ is the platform acceleration, $\dot{\mathbf{r}}$ the platform velocity ($\mathbf{r}$ = position vector), and $\mathbf{f}$ is the vector of measured acceleration (also called specific force). It is assumed that the accelerometers are fixed to the vehicle, which requires a transformation from the vehicle's body frame b to the local level system l. The corresponding rotation matrix $\mathbf{R}$ contains the orientation angles between the two frames. As the platform moves with $\dot{\mathbf{r}}$ with respect to the earth, inertial accelerations arise. These accelerations are taken into account by the last term in (5.93). $\boldsymbol{\omega}_{ie}^l$ and $\boldsymbol{\omega}_{el}^l$ are vectors of angular velocities of the earth's rotation with respect to the inertial frame i and of the platform's rotation with respect to the earth-fixed frame e, respectively.

Operational sea and airborne gravimetry employs modified land gravimeters mounted on a damped two-axes stabilized platform. Stabilization occurs in the

local-level frame by two gyroscope/accelerometer pairs operating in a feed-back mode. For this "scalar" gravimetry, only the *magnitude* of gravity is determined, and (5.93) reduces to

$$g = f_z - \ddot{z} + 2\omega \cos\varphi \sin\alpha \, v + \frac{v^2}{r} . \qquad (5.94)$$

Here, f_z and $\ddot{z}$ are the vertical components of the specific force and the platform acceleration, ω = angular velocity of the earth rotation, φ = geodetic latitude, α = geodetic azimuth, v = platform velocity with respect to the earth, and r = distance to the earth's center. For the static case, (5.94) transforms into the equilibrium conditions of relative gravimetry, cf. [5.4.2].
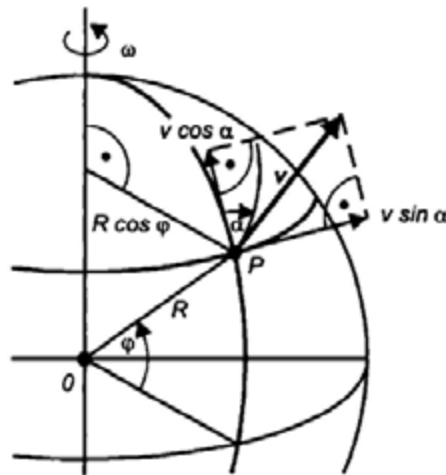

Fig. 5.46. Eötvös effect

The velocity dependent terms on the right side of (5.94) represent the *Eötvös reduction*. From Fig. 5.46, it can be identified as the Coriolis acceleration, which increases (for a west-east directed course) the angular velocity of the earth rotation, and the centrifugal acceleration arising from the platform's angular velocity v/r around the center of the earth. Close to the earth ($r = R = 6371$ km) the Eötvös reduction amounts to

$$\delta g_{E\ddot{o}} = 40 v \cos\varphi \sin\alpha + 0.012 v^2 \; \mu ms^{-2} , \qquad (5.95)$$

with v in km/h. The second term is small for sea gravimetry but attains large values with airborne applications. As velocities can be determined by GPS navigation with an accuracy of 0.05 m/s, the uncertainty of the Eötvös reduction is now less than 10 μms^{-2}.

Instead of using a stabilized platform, the gravity sensor can be rigidly connected to a vehicle. GPS supported inertial navigation systems (INS) are employed with this *strapdown inertial gravimetry*, characterized by a high digital data rate. In the scalar mode, only one approximately-vertical accelerometer is used, while in the vector mode three orthogonally mounted accelerometers determine the specific force vector. According to (5.93), the orientation between the body and the local level frame is needed continuously and is calculated from the output of the INS gyros. Due to the high demands on attitude control, vector gravimetry is still in the experimental stage. Demands are less stringent for scalar gravimetry, especially if the output of the accelerometer triad is used for the determination of the magnitude of gravity (rotation invariant scalar gravimetry), JEKELI (1995), TIMMEN et al. (1998), WEI and SCHWARZ (1998).

The methods for separating *gravity* from *non-gravitational* accelerations depend on the frequency of the accelerations and differ for sea and airborne gravimetry.

High-frequency vibrations can be strongly reduced by the damping of the measuring system. For *sea gravimetry* (stabilized platform), "disturbing" accelerations occur with periods between 2 and 20 s, and they may reach amplitudes of 0.1 g. Due to low ship velocity (10 to 20 km/h) and the nearly constant reference surface (sea level), low-pass filtering sufficiently suppresses the *vertical* accelerations. By averaging the recorded data over time intervals of 1 to 5 minutes, mean gravity values over some 0.1 to 2 km are thus obtained. The effect of *horizontal* accelerations remains small because of the stabilization. Off-leveling effects generally can be neglected (attitude accuracy about 10″). More critical are cross-coupling effects, which occur with horizontal lever spring gravimeters between the horizontal and the vertical component of the disturbing acceleration. They may reach 50 μms^{-2} or more and must be corrected using the horizontal acceleration records. Vertical line gravimeters are free from these errors.

For *airborne gravimetry*, accelerations vary with periods from 1 to 300 s (long-periodic eigenmotion of the airplane) and with amplitudes up to 0.01 g and more. Large airplane velocities (250 to 450 km/h) prevent an effective filtering, and thus with a long filter-length (several minutes) only mean gravity values over some 10 km are obtained. Consequently, accelerations have to be determined independently by differential GPS (carrier phase measurements) through the second time derivative of position or the first time derivative of velocity. Over water and ice areas, radar and laser altimetry can also be employed for height determination. Heights are needed in order to reduce the gravity data to a common reference level, by the free air reduction, cf. [6.5.3].

In 1923, *F.A.Vening-Meinesz* constructed a three-pendulum instrument for gravity measurements in a submerged submarine; world-wide cruises followed until the 1960's. Since that time, *sea gravimeters* (e.g., Askania, LaCoste and Romberg) mounted on gyro-stabilized platforms have operated on board surface vessels, helicopters (1980's), and airplanes (1990's). Since the 1990's,

Fig. 5.47. Gravity sensor Gss30 and gyrostabilized platform KT 30, courtesy Bodenseewerk Geosystem, Überlingen, Germany

these campaigns have also been supported by GPS. A vertical spring is employed in the Bodenseewerk straight-line gravimeter (Fig. 5.47).

Force-balanced *accelerometers*, as developed for inertial navigation, are small and robust with respect to strong dynamics but have less resolution and larger drift rates than conventional land gravimeters. For a linear system, the proof mass is constrained to move in only one direction and maintained at the zero position by an electromagnetic field. The electrical current needed to maintain zero is proportional to the acceleration. Depending on the direction of the sensitive axis, dedicated components of the specific force are measured (BELL and WATTS 1986, NEUMAYER and HEHL 1995), Fig. 5.48. Force-balanced accelerometers are especially suited for use under rough conditions on sea and in air; they have also been employed on board deep sea vessels (COCHRAN et al. 1999). Similar properties are found with *vibrating string* gravimeters. Here, gravity measurements are based on the fact that the vibrational frequency of a string under tension is proportional to the square root of g (BOWIN et al. 1972).
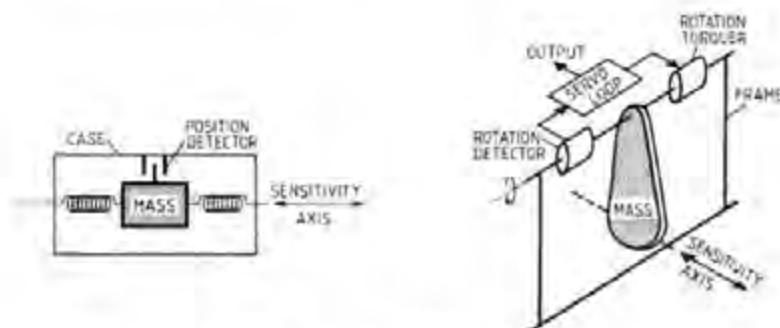

Fig. 5.48. Force-balanced accelerometer principle: translational (suspended mass) system (left) and rotational (pendulum) system (right)

Sea and airborne gravimetric surveys generally are carried out along parallel tracks. Orthogonal tracks serve for control and accuracy improvement by adjustment of the cross-over discrepancies. Sea gravimetry has been concentrated in regions of geological interest and in shelf areas, while recent airborne surveys have covered areas lacking in gravity field information (Greenland, arctic ocean, Antarctica, Switzerland), FORSBERG and BROZENA (1993), KLINGELÈ et al. (1997), Fig. 5.49.

Fig. 5.49. Sea gravimetry profiles (1965-1972), Western Mediterranean Sea, Osservatorio Geofisico Sperimentale, Trieste, after FINETTI and MORELLI (1973)

The *accuracy* of sea and airborne gravity measurements (data recording generally with 1 s average) depends on the survey conditions (sea state, air turbulence, ship and aircraft properties, flight altitude and velocity), on attitude errors, and, for airborne gravimetry, on the separation between gravity and disturbing accelerations. Accuracies of ±5 to 20 μms^{-2} are achieved with sea gravimetry with a resolution of about 1 km along track (track distances 5 to 10 km). Airborne gravimetry generally is carried out at flight heights of a few km, but low speed and elevation (several 100 m) surveys are also performed, especially with helicopters (HAMMER 1983). A resolution of 5 to 10 km is routinely obtained now (helicopter 1 km), with accuracies of ±20 to 50 μms^{-2} (helicopter 5 μms^{-2}). An increase in resolution and accuracy by a factor of two is expected (GUMERT 1995). It must be remembered that the attenuation of the gravity field with height, cf. [3.3.3], prevents a high frequency resolution at high flight altitudes.

5.4.5 Gravity Gradiometry

The *gravity gradient tensor* (3.68) contains local gravity field information, and thus is of interest for high-resolution gravity field determination. It is generally expressed in the local astronomic (local level) system, cf. [3.2.2]. The unit of the components of grad **g** is s^{-2}. In view of the magnitude of the components and the measuring accuracy, they are generally expressed in $10^{-9}\,s^{-2} = ns^{-2}$, traditionally called *Eötvös unit* (E).

A *gravity gradiometer* determines the components of grad **g**, either all or several or linear combinations of them, by exploiting the different reaction of neighboring proof masses to the gravity field. A gradiometer unit consequently consists of two gravity sensors (mostly accelerometers) rigidly connected and generally orientated in the local level system. Taylor expansions of gravity in the two sensors 1 and 2, with respect to the center of mass C of the system, and differences in the output of the sensors (specific force **f**) yields in the stationary mode

$$\mathbf{f}_2 - \mathbf{f}_1 = \left(\operatorname{grad}\mathbf{g}\right)_C \left(\mathbf{r}_2 - \mathbf{r}_1\right)^{\prime}, \tag{5.96}$$

with $\mathbf{r}_1$, $\mathbf{r}_2$ = position vectors of the sensors in the local level system. A gradiometer system is composed of several gradiometer units orientated in different directions in order to derive the corresponding components (Fig. 5.50). Rotation of the gradiometer units in the gravity field provides another means for the determination of different components.

On the earth's surface, *gravimeters* can be used to approximate the components of grad *g* by measuring gravity differences between adjacent stations. The *horizontal* gradient (W_{zx}, W_{zy}) can be derived with a precision of ± 10 ns^{-2} from gravity profiles or area surveys, with station distances of 10 to 100 m (HAMMER
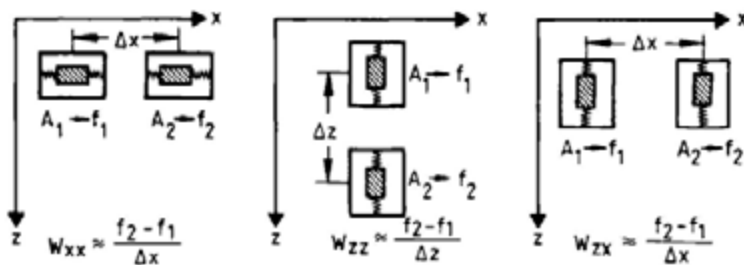

Fig. 5.50. Gravity gradiometer translational system principle with longitudinal (left and center) and transverse (right) constellation

1979). The *vertical* component W_{zz} can be determined with the same precision by repeated relative gravity measurements on tripods, with heights up to 3 m (RÖDER et al. 1985).

The *torsion balance*, developed by *R.v.Eötvös* around 1900, was the first dedicated gravity gradiometer. It consists of two equal masses situated at different heights and rigidly connected by a beam system. At the center of mass the system is suspended by a torsional thread. Equilibrium of the torques acting on the masses is achieved by horizontal rotation, which depends on the components $W_{yy} - W_{xx}$, W_{xy}, W_{xz}, W_{yz}. These quantities and the zero position of the beam are determined by observing the beam direction at five different azimuths. A precision of ±1 to 3 ns^{-2} was obtained (MUELLER et al. 1963). The torsion balance was widely employed in applied geophysics between 1920 and 1940.

Terrestrial gravity-gradiometry in the stationary mode is time consuming and strongly affected by small local mass anomalies. Topographic reductions have to be taken into account even in the immediate surrounding (within 100 m), which limits the application to flat or moderate hilly areas.

Measurements on *moving platforms* allow rapid data collection (e.g., in a 1 s rate or more). With airborne and satellite applications, cf. [5.2.7], topographic effects are significantly reduced. In the kinematic mode, the gradiometer system is mounted on a gyro-stabilized platform. The effects of the platform's rotation about the earth have to be taken into account, cf. [5.4.4]. Due to differencing, non-gravitational forces of linear type cancel with this method.

A gravity gradiometer survey system for automobile and airplane applications has been developed at Bell Aerospace, with a precision of 10 ns^{-2} and a resolution of a few km (JEKELI 1988a, VASCO and TAYLOR 1991). It consists of three gradiometer units, each equipped with two accelerometer pairs, mounted orthogonally on a slowly rotating disk (Fig. 5.51). The gradiometer units are combined under different orientation on a gyro-stabilized platform. Superconducting technology may greatly reduce vibrational and long-term effects and provide a precision of 1 ns^{-2}. Airborne or shipborne surveys could then resolve the gravity field down to 1 km or less and by integration give an accuracy of 1 μms^{-2} for gravity (PAIK et al. 1997).
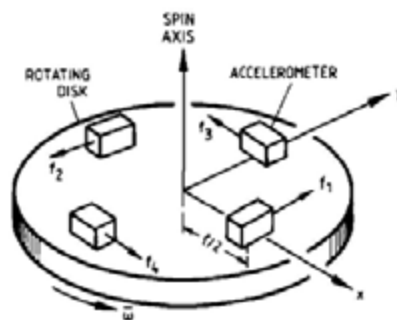

Fig. 5.51. Rotating gravity gradiometer principle, after JEKELI (1988a)

5.4.6 Continuous Gravity Measurements

Continuous gravity records contain information on earth tides and ocean tidal loading, free oscillations and the nearly diurnal free wobble of the earth, inner core translation, polar motion, and volcanic and earthquake activity (ZÜRN et al. 1997, CROSSLEY et al. 1999), cf. [8.3.5]. These effects occur at time scales between seconds and several years and have amplitudes of about 1 to 1000 (tides) nms^{-2}. Hence, a *recording gravimeter* should provide a resolution of 0.01 to 1 nms^{-2} and a high stability with time (low drift rates). It should be time-controlled within ±10 ms. In order to reduce environmental effects (temperature changes, microseismicity, local inclinations), recording gravimeters generally are installed at underground sites (basement, tunnel).

Recording gravimeters operate in an electronic feedback mode, cf.[5.4.2], over a limited measuring range, e.g., 10 μms^{-2}. The voltage output is proportional to gravity and first undergoes an analogue filtering in order to reduce the high-frequency noise. It is then digitized by an A/D converter. Digital filtering delivers a data set (1 to 10 s samples), which is stored on a PC. Further numerical filtering and data reduction may be appropriate as well as the reduction of spikes due to earthquakes and the interpolation of data gaps (WENZEL 1996). An analogue output offers a convenient on-line control of the data acquisition.

Spring-type and superconducting gravimeters are used for gravity recording (MELCHIOR 1983).

Elastic spring gravimeters can be employed if supplemented with a low-pass filter, a recording unit, and a quartz clock. Special earth tide gravimeters have also been developed and are characterized by long-term stability (e.g., by a double thermostat). Recent land gravimeters offer the option of an earth tides mode through increased sensitivity, large memory, and computer-controlled remote operation. The long-term drift of these instruments has to be removed by filtering. Consequently, only short-period effects (e.g., diurnal and higher-frequency tides) can be determined, at a noise level of a few 0.1 to 1 nms^{-2}.

For the *superconducting* gravimeter, the gravity acting on the proof mass (Nb sphere) is compensated by a magnetic counterforce (Fig. 5.52). The magnetic field is generated by superconducting coils and thus is extremely stable with time. The position of the mass is monitored by a capacitive detector and held fixed by a feedback coil. Liquid helium provides the superconducting state at a temperature of 4.2 K. The measuring system is kept in an insulating Dewar vessel (Fig. 5.53). For long-term remote operation, a reduced size option is available with no need for liquid helium refilling (GOODKIND 1991, RICHTER
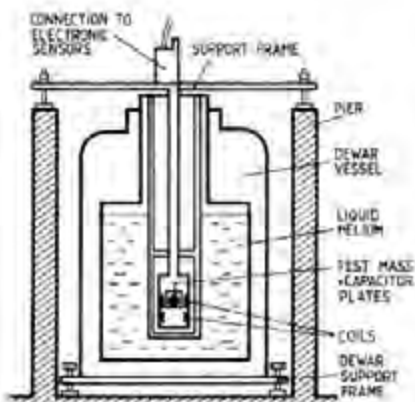
Fig. 5.52. Superconducting gravimeter principle, after GWR-Instruments information and RICHTER (1987)

Fig. 5.53. RSG remote controlled tidal superconducting gravimeter view, courtesy of GWR-Instruments Inc., San Diego, CA, U.S.A.

1995, WARBURTON and BRINTON 1995). The instrumental drift of a superconducting gravimeter is very small (on the order of 10 nms^{-2}/a), and may be modeled by an exponential function. Absolute gravity measurements can be used to control the drift at longer time intervals. From tidal analysis, a noise level between 0.01 to 0.1 nms^{-2} (short periods) and 1 to 10 nms^{-2} (very long time intervals) has been found (HINDERER et al. 1994, NEUMEYER and DITTFELD 1997).

Calibration of a recording gravimeter is performed by relative and absolute methods. A *relative* calibration is realized by parallel registration with a "calibrated" gravimeter or by recording on a station with well-known tidal parameters. *Absolute* calibration is generally performed on a vertical-gravimeter calibration line. Laboratory calibration procedures include the artificial periodic acceleration on a vertically-oscillating platform, the controlled vertical displacement of large external masses, and the parallel registration with an absolute gravimeter. Accuracies of a few 0.1 % are achieved (RICHTER et al. 1995, FRANCIS 1997). The instrumental phase shift is determined by recording the gravimeter's response to a defined impulse.

The gravity signal is strongly correlated with *atmospheric pressure*. A linear regression with local air pressure (between –2.5 and -3.5 nms^{-2}/hPa) reduces the main part of this effect. More refined models are available which also take the air pressure around the station and the elastic response of the earth's crust into account (MERRIAM 1992). Variations of *groundwater* level and *precipitation* also affect the gravity record (order of some 10 nms^{-2}) but are difficult to model (VIRTANEN 2000).

Longer (several months and more) gravity records can be subjected to a *tidal analysis*. It is based on the spectral decomposition of the observed signal into a number of partial tides. By comparing the observations (hourly samples, atmospheric pressure effects reduced) with the gravimetric tides for a rigid earth, cf. [3.5.2], deviations in amplitude and phase are found, which depend on the earth's elastic response to the tidal forces. For the partial tide *i*, this is expressed by the *amplitude factor*

$$\delta_i = A_i(\text{obs}) : A_i(\text{theor}), \tag{5.97}$$

with A_i = observed resp. calculated (solid earth) amplitude, and the *phase shift*

$$\Delta\Phi_i = \Phi_i(\text{obs}) - \Phi_i(\text{theor}), \tag{5.98}$$

with Φ_i = observed resp. calculated phase. The observation equation for a least squares *spectral analysis* then reads

$$l(t) = \sum_{i=1}^{n} \delta_i A_i(\text{theor}) \cos(\omega_i t + \Phi_i(\text{theor}) + \Delta\Phi_i), \tag{5.99}$$

with $l(t)$ = recorded gravity value at time *t*, and ω_i = circular frequency for the partial tide *i* (WENZEL 1976, 1997).

Elastic spring gravimeters allow the determination of 10 to 20 partial tides (mainly diurnal, semidiurnal, terdiurnal), with an observation time of 4 to 6 months. Superconducting gravimeters can resolve up to 40 tides (including semiannual and annual) by registration over several years. The gravimetric factor for polar motion has also been derived from long-term series. As an example, a 158 days registration with a LaCoste and Romberg feedback gravimeter at Hannover ($\varphi = 52.387°$N, $\lambda = 9.713°$E, $H = 50$ m) yielded for the lunar diurnal tide O1

$$\delta(\text{O1}) = 1.151 \pm 0.001, \ \Delta\Phi(\text{O1}) = 0.16° \pm 0.08°$$

and for the semidiurnal tide M2

$$\delta(\text{M2}) = 1.188 \pm 0.0005, \ \Delta\Phi(\text{M2}) = 1.70° \pm 0.03°$$

(TIMMEN and WENZEL 1994a). The factor for O1 is close to the observed global value 1.155, while the M2 result differs due to ocean load and attraction, cf. [8.3.5].

A global data bank for tidal gravity measurements (more than 300 stations) is maintained at the International Center for Earth Tides in Brussels (MELCHIOR 1994).

5.5   Terrestrial Geodetic Measurements

Terrestrial geodetic measurements are carried out by directly observing geometric quantities between points on the earth's surface. The majority of the observations refer to the gravity field of the earth through orientation in the local astronomic systems. The measurement of horizontal and zenith angles [5.5.1] and distances [5.5.2] allows relative positioning, where combined instruments (total stations) are generally used. Very precise height differences are provided by leveling [5.5.3]. Strain and tilt measurements serve for detecting local changes of distances and inclination with time [5.5.4].

Due to the high accuracy and economy of satellite-based measurement techniques, terrestrial geodetic measurements are used primarily for interpolating satellite-derived results or in areas where satellite methods fail or need terrestrial support (underground and underwater positioning, forests, urban areas, engineering surveys, local geodynamics).

Inertial surveying provides kinematic three-dimensional positioning by evaluating accelerometer measurements on a moving gyro-stabilized platform. The influence of the gravity field has to be taken into account by reductions. The method is mainly used for station densification in control networks, especially for mapping purposes. The accuracy achieved depends on distance and is about $\pm (0.1 \text{ m} + 10 \times 10^{-6} \text{ s})$, SCHWARZ (1986), JEKELI (2001).

Terrestrial methods are treated in textbooks on surveying, e.g., KAHMEN and FAIG (1988), BANNISTER et al. (1998). Classical surveying instruments are described by DEUMLICH (1988), while the actual state is dealt with in JOECKEL und STOBER (1999), and DEUMLICH and STAIGER (1999). KAHMEN (1978) and SCHLEMMER (1996) concentrate on the fundamentals of electronics employed in geodetic instruments, while Brunner (1984b) deals with the effects of atmospheric refraction.

5.5.1 Horizontal and Vertical Angle Measurements

The *horizontal angle* is defined as the angle measured in the horizontal plane of the local astronomic system between two vertical planes. It is formed by the difference in horizontal *directions* to the target points which define the vertical planes. The *vertical angle* is the angle measured in the vertical plane between the horizontal plane and the direction to the target point. Generally the *zenith angle*, the complement to 90°, is introduced instead of the vertical angle, cf. [2.6.2].
A *theodolite* is used for measuring horizontal and vertical angles. The principal components of this instrument are a horizontal and a vertical circle with graduation, a telescope capable of being rotated about the vertical and the horizontal axes, and a mechanism for reading the circles. In order to orientate

the theodolite with respect to the plumb line direction, it is equipped with spirit or electronic levels.

Regarding the reading of the circle graduation, we distinguish between optical and electronic or digital theodolites.

Optical theodolites of highest precision were developed since the second half of the 18th century and were then used until the 1960's for first order triangulation (with the Kern DKM3 and the Wild T3 theodolites being the latest developments) at station distances of 30 to 60 km. They were characterized by very stable construction and circle diameters of 100 to 250 mm. Circle graduation errors were less than 0.5″, and reading accuracy reached 0.1″ by using a coincidence microscope with micrometer screw. The lens aperture of the telescope was 60 to 70 mm and the magnification 30 to 40 or more. Standard deviations of ±0.2″ to 0.4″ have been achieved for an adjusted horizontal direction, cf. [7.1.1].

Today, horizontal angles are measured only over maximum distances of a few km for densification surveys or dedicated local networks. *Electronic theodolites* (lens aperture 40 to 45 mm, magnification of 30 or more) are available for this purpose and have superseded the optical analogue instruments (Fig. 5.54). Generally, the electronic theodolite is combined with a distance meter to produce a total station, cf. [5.5.2].

The horizontal and vertical circles of an electronic theodolite are either coded or carry an incremental graduation. Reading is microprocessor-controlled and performed by optical-electronic scanning and subsequent interpolation (electronic micrometer). Electronic levels and a dual-axes compensator serve for leveling the instrument; a residual tilt correction may also be applied automatically. Collimation and horizontal axis errors are either eliminated by measurement at both positions of the telescope or corrected internally. An accuracy of ±0.5″ to 1″ is obtained for a direction observed in two positions.

Gyrotheodolites have been developed for the determination of astronomic azimuths by combining a theodolite with a gyroscope. The principle of the gyroscope is based on the fact that a rapidly rotating gyro with horizontal spin axis swings into the north direction due to the mutual effects of the gyro's spin, the earth's gravity, and the earth's rotation. An accuracy of ±3″ can be obtained from 20 individual measurements. Gyrotheodolites are employed primarily for mining and tunnel surveys, Fig. 5.55.

For the measurement of *zenith angles,* the theodolite is equipped with a reading index for the vertical circle. By leveling the index either manually (spirit level) or automatically, the local vertical is realized with an accuracy of a few 0.1″. With electronic theodolites (circle diameter about 70 mm, electronic levels with 2″/2 mm) an accuracy of ±0.5″ to 2″ is obtained for an observed zenith angle.

Fig. 5.54. Electronic theodolite
(Leica T 1800), courtesy Leica
Geosystems AG, Heerbrugg,
Switzerland

Fig. 5.55. Gyrotheodolite (Gyromat
2000), courtesy Deutsche Montan
Technologie (DMT), Essen, Germany

Further processing of zenith angles, cf. [6.4.2], requires the consideration of errors due to vertical refraction.

The *refraction angle* depends on the coefficient of refraction (5.11) and thereby on the meteorological conditions along the path of light, particularly the vertical gradient of temperature, see (5.19). Generally, the refraction angle is derived from meteorological data taken at the endpoints of the observation line. This easily leads to errors of a few arcsec and more for distances larger than a few km. According to (5.11b), the effect of this error on the height difference increases with the square of the distance and reaches the order of a few decimeters over a few km. The error of the observed zenith angle, on the other hand, only propagates with distance and thus remains at the order of a few cm.

An approximately symmetric behavior of refraction is to be expected for *simultaneous* observations at the endpoints, especially with cloudy weather and prior to the isothermal conditions of the evening and if the ray of light is more than 15 to 20 m above the ground. The uncertainty of the refraction angle remains less than 1″ for distances below 10 to 25 km. This has led to the method of observing reciprocal simultaneous-zenith-angles.

For the direct determination of the refraction angle, the *dispersion* of the light may be utilized, cf. [5.1.1]. As the light path at shorter wavelengths has a greater curvature than at longer ones, the use of two different wavelengths causes a difference between the two angles of refraction at the target point. This dispersion angle depends on the effective coefficients of refraction along the light path. Neglecting the small influence of water vapor pressure, it is proportional to the refraction angle but about two orders of magnitude smaller. The main error source of the measurement of the dispersion angle is atmospheric turbulence. Experiments at distances less than 20 km gave uncertainties of ±1 to 2″ for the refraction angle. An eventual application of

dispersometers probably will be restricted to distances of a few km (WILLIAMS and KAHMEN 1984).

5.5.2 Distance Measurements, Total Stations

Terrestrial distances represent important geometric quantities for positioning. They provide geometric relations between points and establish the scale of classical geodetic networks.

Until about 1960, the scale of triangulation networks, constructed from angle measurements, was derived from *base lines* having lengths of 5 to 10 km. Measuring rods and, since 1900, wires or tapes served to measure the base line length. With the *Jäderin* (1880) method, freely hanging invar (NiFe alloy) wires 24 m in length were used, characterized by a small coefficient of thermal expansion. The relative accuracy of the more recent base lines amounts to 10^{-6}, which corresponds to 1 mm/ 1 km.

For field calibrations of wires and tapes, several international *calibration lines* were established by interferometric methods. Starting from the length of a standard meter, the *Väisälä* light interference comparator provided an optical multiplication up to base line lengths of 864 m (relative accuracy $\pm 10^{-7}$).

Electromagnetic distance measurements started at the end of the 1940's. They either use light waves ($\lambda = 0.4$ to $0.8\ \mu m$) and the near infrared (up to $\lambda = 1\ \mu m$) or microwaves ($\lambda = 1$ to 10 cm) as carriers of the measuring signal (RUEGER 1997, JOECKEL und STOBER 1999). The travel time of the signal serves as a measure for the distance.

Microwaves are hardly absorbed by the atmosphere and allow the measurement of large distances (50 km and more) even under unfavorable weather conditions. The effect of humidity on refraction, on the other hand, is great and may significantly deteriorate the results. Distances measured by *light waves* are about one order of magnitude more accurate, but their use may be limited by visibility disturbances (haze, fog), cf. [5.1.2].

Time measurement is performed by the pulse or the phase comparison method.

For the *pulse method*, the transmitter emits a pulse which is reflected at the target and observed at the receiver. An electronic timer measures the time Δt that the signal requires to travel forth and back along the distance s. We have

$$s = \frac{c}{2}\Delta t\ , \tag{5.100}$$

where we assume that refraction effects have been taken into account by corresponding corrections, cf. [5.1.1]. If the uncertainty in distance is to remain less than 5 mm, the time of propagation must be obtained to an accuracy of ±0.03 ns. This high accuracy demand can be fulfilled by short (a few ns) laser pulses, electronic counting with a high frequency oscillator, and averaging the results of a large amount (e.g., 1000) of individual measurements, cf. also [5.2.6].

For the *phase comparison method*, a high-frequency carrier wave is sent out by the transmitter and modulated continuously (amplitude or frequency modulation), with modulation frequencies between about 10 and 100 MHz. The corresponding half wavelength (because of the double distance traveled by the signal) serves as a "yard stick" (about 1 to 10 m) in surveying the distance (Fig. 5.56). The phase shift found between the transmitted and the received signal represents the residual part of the distance above an integral number of complete wavelengths.
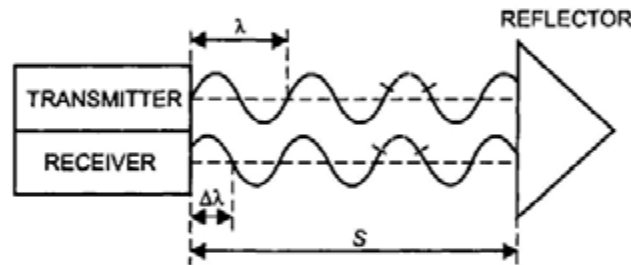
Fig. 5.56. Phase comparison method principle

Travel time Δt and phase shift $\Delta \Phi$ are related through

$$\Delta t = \frac{N + \Delta \varphi / 2\pi}{f},$$

(5.101)

with N = number of complete periods, and the modulation frequency

$$f = \frac{v}{\lambda} = \frac{c}{n\lambda},$$

(5.102)

n = refractive index, cf. [5.1.1]. Substituting (5.101) and (5.102) into (5.100), and assuming that refraction is corrected separately, delivers the distance

$$s = \frac{\lambda}{2}\left(N + \frac{\Delta\varphi}{2\pi}\right). \tag{5.103}$$

With the residual part of a wavelength

$$\Delta\lambda = \frac{\Delta\varphi}{2\pi}\lambda, \tag{5.104}$$

the distance can also be expressed by

$$s = N\frac{\lambda}{2} + \frac{\Delta\lambda}{2}. \tag{5.105}$$

Hence, a distance measuring unit consists of an oscillator and transmitter, a receiver, a phase meter, and a microprocessor. With a digital phase detector, the measuring process can be fully automatized; the resolution achievable is 10^{-3} to 10^{-4}, which corresponds to a mm-precision. The number N is determined automatically by applying several slightly different modulation frequencies generated by frequency division.

Terrestrial *microwave* distance measurements started with the development of the tellurometer by *T. L. Wadley* (1956). Here, the master station emitted a modulated (modulation frequencies between 7.5 and 150 MHz) carrier wave (λ = 8 mm to 10 cm), which was retransmitted from the active transponder (receiver and transmitter). Measurement of ranges up to 70 km and more were obtained. The accuracy strongly depended on refraction uncertainties and could reach

$$\pm\left(10...15\,\mathrm{mm} + 3\times10^{-6}\,s\right).$$

Electro-optical distance measurements trace back to the first geodimeter developed by *E. Bergstrand* (1948). Long-range distance meters used laser light (He-Ne gas laser) with modulation frequencies between 15 and 50 MHz and were able to measure distances up to 60 km on clear days, with an accuracy of

$$\pm\left(1...5\,\mathrm{mm} + 1...2\times10^{-6}\,s\right).$$

Microwave and long-range electro-optical distance measurements have been carried out extensively from the 1950's to the 1970's. The measurements were primarily for establishing first order control networks and for strengthening existing horizontal control, cf. [7.1.1]. Long-range distance meters have become obsolete, as large-scale positioning is carried out nowadays almost exclusively by satellite methods.

Today, terrestrial distance measurements are restricted to ranges of a few km and about 10 km maximum. Visible light and (mostly) near infrared are used, either in the pulse or (mainly) in the phase comparison method. Pure distance

meters are available for special applications (partly suitable to be mounted on a theodolite), Fig. 5.57, but generally electronic *total stations* are employed, measuring distances, horizontal directions and zenith angles simultaneously.

Fig. 5.57. Pulsed infrared laser distance meter (Leica Di 3000S), courtesy Leica Geosystems AG, Heerbrugg, Switzerland

The measuring range of a distance measuring unit depends on the number of prisms posted on the target station. With pulsed infrared laser light (gallium arsenide diode), distances up to 5 km (1 prism) or 10 to 20 km (multiple prisms) can be observed with an *accuracy* of

$$\pm\left(3...5\,\text{mm} + 1 \times 10^{-6}\,s\right)$$

and within a measuring time of a few seconds.

With the phase comparison method, using modulated infrared light, distances up to 2 to 3 km can be measured with one prism with an accuracy of

$$\pm\left(1...2\,\text{mm} + 1...2 \times 10^{-6}\,s\right).$$

When integrated into a total station, the microprocessor-controlled operation also includes the (partly) automatic searching and pointing of the target, leveling by a two-axes compensator (range 5′, precision ±0.3″), measurement of horizontal direction and zenith angle, data storage in the internal memory, refraction corrections with standard or actual meteorological data, transformation from polar to local Cartesian coordinates, and a graphic display (FEIST et al. 1998), Figs. 5.58, 5.59.

The *calibration* of electronic distance meters includes the control of the modulation frequency by a frequency meter and the determination of the instrumental constants (zero point correction and possible cyclic errors) on a comparator or on a short (about 1 km) calibration line. Calibration lines are usually partitioned into several sections and determined by laser interferometry

Fig. 5.58. Total station (Trimble TTS 500), courtesy Trimble Navigation Ltd., Sunnyvale, CA, U.S.A.

Fig. 5.59. Total station (Geodimeter System 600), courtesy Spectra Precision AB, Danderyd, Sweden

or with a short-range distance meter of high precision.

The error budget of distance measurements contains a constant part that depends on uncertainties in timing or phase measurement and on the zero point stability. The distance dependent part is determined by errors of the modulation frequency and by refraction effects.

The *meteorological parameters* generally are measured only at the instrument and at the target point, and the arithmetic mean is introduced with the refraction reduction. This value may not be representative for the entire distance, and thus a limiting factor for precise distance measurements is set (ILIFFE and DODSON 1987). Measurements conducted under the same atmospheric conditions may be highly correlated, but this correlation can be significantly reduced if observations are carried out under different conditions (HÖPCKE 1965).

Special distance meters have been developed in order to measure shorter base lines with very *high precision* (calibration lines, local geodynamic control). These developments are based either on instrumental refinements and better determination of the refraction effects or on the use of two or three different wavelengths.

The Mekometer ME 5000 uses a He-Ne laser as a light source, with a polarization modulated wavelength of 0.6 m (FROOME 1971, MEIER and LÖSER 1986), Fig. 5.60. The modulation frequency is automatically adjusted such that the distance becomes an integer number of wavelengths and is measured with a frequency meter. The Geomensor (COM-RAD, England) uses a Xenon flash tube as light source (SCHERER 1985). Meteorological data are provided by a special sensor unit. Distances of 5 km can be measured with one prism , and a precision of

$$\pm\left(0.1...0.2\,\text{mm} + 0.2 \times 10^{-6}\,s\right)$$

may be obtained. However, refraction effects may increase the distance-dependent contribution to $0.5 \times 10^{-6}\,s$ (DODSON and FLEMING 1988).

Fig. 5.60. Distance meter Mekometer ME 5000, courtesy Leica Geosystems AG, Heerbrugg, Switzerland

If the distance is measured with different wavelengths, the effect of *dispersion* can be exploited, cf. [5.1.1]. For light waves, the difference in the distances obtained with "blue" and "red" light particularly depends on the influence of temperature and atmospheric pressure on the index of refraction. If an additional microwave measurement is performed, the effect of humidity is then reflected mainly in the deviation of the microwave results from the light results. From (5.14), (5.15), (5.17), and (5.18), a relation between the geometric length of the path and the differences between the distances obtained with different light waves and a microwave can be established. A resolution of a few 0.01 mm is required at these small differences in order to obtain mm-accuracy for the distance. The terrameter has been developed as a two-wave instrument (He-Cd laser 441.6 nm, He-Ne laser 632.8 nm) to measure distances with a precision of ±0.1 mm (LANGBEIN et al. 1987). By adding a microwave ($\lambda = 3$ cm) and employing a light modulation frequency of 3 GHz, the relative accuracy of this method could be increased to 10^{-7} and better (HUGGETT and SLATER 1975).

The *refraction corrections* which have to be applied to the observed distances before further processing can be split up into three parts (HÖPCKE 1966). The distance $\overline{s}_0$ read on the instrument is based on a standard value n_0 for the refractive index calculated from standard temperature and air pressure (5.17). If a more realistic value $\overline{n}$ is available from local meteorological measurements, the relation

$$\overline{s}\,\overline{n} = \overline{s}_0 n_0 \qquad\qquad (5.106)$$

provides a first *velocity correction*

$$k_n = \overline{s}\left(n_0 - \overline{n}\right). \tag{5.107}$$

As seen from (5.20a), the radius of the light curve differs from the earth's radius ($r \approx 8\ R$). Hence, the light passes through atmospheric layers with a greater density and refractive index than estimated by the mean value $\overline{n}$, calculated from the endpoint data. With (5.10) and the coefficient of refraction $k = 0.13$, we obtain

$$\frac{dn}{dh} = -\frac{k}{R} = -20 \times 10^{-6}/\text{km}, \tag{5.108a}$$

which can be used to derive a second velocity correction

$$k_{\Delta n} = -\left(k - k^2\right)\frac{\overline{s}^3}{12R^2}. \tag{5.108b}$$

This correction is less than 1 mm over a distance of 15 km and can be neglected generally.

For three-dimensional computations, the *chord distance s* is required. By assuming a spherical arc with radius r we have

$$s = 2r \sin\frac{\overline{s}}{2r},$$

or after a series expansion

$$s = 2r\left(\frac{\overline{s}}{2r} - \frac{1}{6}\left(\frac{\overline{s}}{2r}\right)^3 + \ldots\right).$$

Introducing (5.10) yields the curvature reduction

$$k_r = -k^2\frac{\overline{s}^3}{24R^2}, \tag{5.109}$$

which is part of the reduction formula (5.6). This reduction is less than 0.1 mm for a distance of 15 km and can be neglected. By adding (5.107) to (5.109), we obtain the total reduction from the observed distance to the chord:

$$s - \bar{s}_0 = \bar{s}\left(n_0 - \bar{n}\right) - \frac{2k - k^2}{24R^2}\bar{s}^3 . \qquad (5.110)$$

The reduction from the chord distance to the length of the normal section and the geodesics on the ellipsoid will be given in [6.3.2].
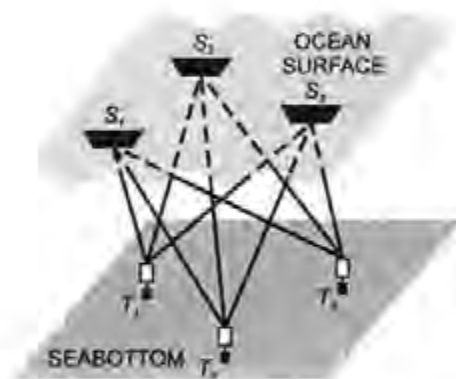

Fig. 5.61. Acoustic positioning on the sea bottom (T = transponder, S = transmitter)

Control points have been established locally on the *ocean floor* mainly for geodynamic investigations (e.g., sea floor spreading at active ridge zones). Control points are usually arranged in arrays of 3 to 4 stations, with station separation of 5 to 10 km. *Acoustic waves* (velocity about 1500 m/s in water) are used exclusively for relative positioning, with acoustic transponders powered by batteries or sources of nuclear energy. Acoustic signals (5 to 20 kHz) are emitted from a ship-borne transducer and sent back by the transponders. The slant range is calculated from the propagation time of the signal traveling forth and back. Spatial trilateration then provides the relative positions (RINNER 1977), Fig. 5.61. With microsecond timing, the accuracy of the acoustic wave method is determined by the sound velocity, which depends on temperature, salinity, and water pressure. cm-accuracy can be achieved over a few km (CHADWELL et al. 1998). The relation to the global reference system is established by GPS-positioning on board of the surface vessel.

5.5.3 Leveling

In *geometric leveling*, differences in height are determined using horizontal lines of sight between points in close proximity to each other. Leveling is conducted with a *leveling instrument* (level) and two vertically posted leveling rods (Fig. 5.62). The leveled height difference δn between the rods is given by the difference between the backsight (b) and the foresight (f) reading:

$$\delta n = b - f . \tag{5.111}$$

The *leveling instrument* consists primarily of a measuring telescope capable of rotation about the vertical axis. The line of sight is brought into the horizontal either by a coincidence bubble in conjunction with a tilting screw or, for most modern instruments, automatically by a compensator that is comprised mainly of a gravity pendulum (first introduced with the ZEISS Ni2 in 1950). A setting accuracy of ±0.1″ to 0.4″ is achieved by both methods. The use of a compensator increases the speed of leveling, but spirit levels offer advantages if high-frequency oscillations (traffic, wind) occur, due to better damping.
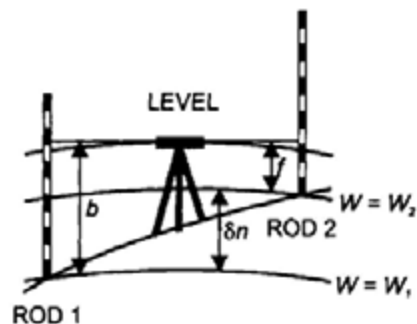

Fig. 5.62. Geometric leveling principle

High-precision levels employ telescopes with an aperture of 40 to 50 mm and a magnification of 30 to 40. Distances between the level and the rods are kept within 30 to 40 m. Setting-up the instrument in the middle of two subsequent rod positions is typical, as it eliminates errors due to non-parallelism of the collimation and the bubble axes as well as symmetric refraction effects. In the *analogue* mode, the leveling rods carry two graduation lines on invar tape, displaced against each other and numbered differently in order to detect reading errors. The line of sight is adjusted to the closest graduation mark by means of a parallel plate mounted in front of the objective's lens. The amount of the displacement is measured by a micrometer (Fig. 5.63).

Digital levels were introduced with the Wild NA 2000 (INGENSAND 1990). They are used in connection with invar staffs that carry a binary code (Fig. 5.64). A code section around the horizontal sight is projected on a CCD sensor in the image plane of the telescope. A subsequent processing of the image by a microprocessor includes electronic scanning (A/D conversion) and correlation with a digital reference signal, whereby the automatically measured distance has to be taken into account (MEIER-HIRMER 1997, RUEGER and BRUNNER 2000).

Fig. 5.63. Bubble level (Wild/ Leica N3), courtesy Leica Geosystems AG, Heerbrugg, Switzerland

Fig. 5.64. Digital level (Zeiss DiNi 12), courtesy Zeiss/Spectra Precision AB, Danderyd, Sweden

In order to transfer heights over larger distances, the individual leveled differences are summed. For one set-up, the non-parallelism of the level surfaces may be neglected (quasi-differential method). The observed difference δn then corresponds to the height difference of the level surfaces passing through the rod sites. Summing the individual differences between two bench marks P_1 and P_2 yields the "raw" leveled height difference

$$\Delta n_{1,2} = \sum_{1}^{2} \delta n . \tag{5.112}$$

At longer distances, the effect of the non-parallelism may reach the cm-order of magnitude and more, cf. [3.2.1]. Hence Δn depends on the path taken and does not provide a unique height in any height system. A unique height determination can be achieved only by considering gravity g, that is by referring to potential differences ΔW. According to (3.52) we have

$$\Delta W_{1,2} = W_2 - W_1 = -\int_{1}^{2} g \, dn \approx -\sum_{1}^{2} g \, \delta n . \tag{5.113}$$

Thus potential differences can be determined without any hypothesis from leveling and surface gravity. In order to obtain height differences in any specific height system from the raw leveling results, gravity reductions have to be applied, cf. [6.4.1].

The *accuracy* of precise leveling depends on many influences. Some of the leveling errors behave in a random manner and propagate with the square root of the number of individual set-ups. Other errors are of systematic type and may propagate with distance in a less favorable way. Hence particular attention must

be afforded to reduce them, by instrumental measures and modeling or by employing dedicated measurement methods.

Important *error sources* include (KUKKAMÄKI 1980):

- Misleveling of the instrument. This random type error is of the order of a few 0.1″ and results in an error of a few 0.01 mm for an individual height difference. Systematic effects arise from a residual adjustment error of the bubble or an imperfect operation of the compensator and cause an "obliquity of the horizon". These effects cancel when measuring in two opposite positions of the compensator.
- Magnetic effects on the compensator. These are kept small generally, but a regular control is advisable.
- Rod graduation errors. This includes a "mean" scale error, errors of the individual graduation marks, and effects of thermal expansion. Routine calibrations deliver the corresponding corrections, with a remaining random part less than 5 to 10 μm (SCHLEMMER 1984).
- Rod inclination errors. These errors can be kept sufficiently small by properly adjusting the rod bubble and carefully holding the rod in the vertical position.
- Vertical refraction. This depends mainly on the vertical temperature gradient, cf. [5.1.2]. The irregular part (shimmer) acts as a random error (±0.01 mm under cloudy skies). Systematic influences occur particularly with measurements made in terrain with steep slopes or those made close to the ground. The effects may reach 0.01 to 0.1 mm per 1 m height difference. They can be modeled in part by a refraction correction, with the vertical temperature gradient as a function of height (KUKKAMÄKI 1938). More refined models are available, which include other meteorological and hydrological data, terrain slope, and orientation with respect to the sun (ANGUS-LEPPAN 1984).
- Vertical movements of the instrument and the rods. These errors depend on the stability of the ground and on the manner in which the instrument is set up; movements of 0.01 to 0.1 mm per station are possible. Movements proportional to time are cancelled by an observation succession b_{I}, f_{I}, f_{II}, b_{II} (I, II = scales of the rod). In addition, from the mean of forward and backward leveling runs, those rod movements are eliminated which are proportional to time and which occur while the instrument is brought to the next station.
- Earth tides effects. The tidal effect on the plumb line causes periodic inclinations of the line of sight. The inclinations can be modeled by the horizontal tidal component acting in the azimuth of the leveling line. Starting from (3.119) and taking the elasticity of the earth into account, the tidal reduction for the moon reads

$$\delta_{t(m)} = 0.06 \sin 2\psi_m \cos(\alpha_m - \alpha) s \ \mathrm{mm/km} \ . \tag{5.114}$$

Here, α_m and α are the azimuths of the moon and of the leveling line, and ψ_m the geocentric central angle between the directions to the moon and the computation point. s is the length of the leveling line in km. The same equation is valid for the sun, with an effect of 46% of that for the moon (KUKKAMÄKI 1949).

In order to eliminate or reduce systematic errors, precise leveling is always carried out with equal back and foresights ("leveling from the middle") at less

than 50 m distance. Observations should be performed during cloudy weather, preferably in the morning and in the evening hours. Line of sights very close to the ground (0.5 m or less) should be avoided. Leveling is always conducted twice, in opposing directions and possibly under different meteorological conditions. For a 1 km double-run leveling, one can attain a standard deviation of 0.2 to 1.0 mm. Residual systematic effects (especially those due to local vertical movements and refraction) may cause small correlations between the height differences obtained from the forward and backward runs and affect the results of large network adjustments (LUCHT 1972).

The time needed for precise leveling can be significantly reduced by employing a *motorized* procedure, whereby the instrument and the rods are carried in and operated from an automobile. This mode also reduces time dependent errors and partially eliminates asymmetric refraction effects, as the line of sight is more remote from the ground (PESCHEL 1974, BECKER 1987).

For leveling across broad *waterways* and inlets of the *sea*, several methods have been developed:

- In *reciprocal leveling*, approximately horizontal sights to specially designed targets are taken simultaneously with a precise level from both sides of the waterway. For longer series of observations including a change of the instruments, height differences over 1 to 2 km can be determined with a precision of ±1 to 2 mm (JELSTRUP 1955, KAKKURI 1966).
- *Hydrostatic leveling* is based on the principle of communicating tubes. A hose filled with water (free of air bubbles, uniform temperature) is laid between the shores of the watercourse to be bridged. The water level observed at the vertical ends of the hose belong to the same level surface (GRABOWSKI 1987). The method may be applied for ranges up to 20 km (Fehmarn-Belt/Baltic Sea) and delivers mm-precision (ANDERSEN 1992). In the Netherlands, it is used in an operational mode (WAALEWIJN 1964).
- In *hydrodynamic leveling* (geostrophic leveling), the height is transferred over the waterway utilizing water level records, which have to be reduced for the effects of sea surface topography, cf. [3.4.2]. This implies the development of a hydrodynamic model, which takes water velocity, wind drag, water depth and bottom friction, atmospheric pressure, water density, and gravity and Coriolis force into account. The method has been applied, for instance, for a height transfer over the British channel (70 km) and over the Fehmarn-Belt (CARTWRIGHT and CREASE 1963, WÜBBELMANN 1992) with a precision of about 1 cm.

5.5.4 Tilt and Strain Measurements

Tilt and strain observed on the surface of the earth indicate the response of the earth's crust to external and internal forces such as earth tides, tectonic processes, and seismic and volcanic activities. Tilt and strain are dimensionless quantities and are given in radian or arcsec and extension (positive sign) or compression per distance, respectively.

Over time intervals of years to decades, long-term tilt and strain can be determined from repeated observations of geodetic control networks using satellite and terrestrial techniques, cf. [8.3.3]. Tiltmeters and strainmeters (also called extensometers), on the other hand, have been developed in order to monitor continuously local deformations (MELCHIOR 1983, AGNEW 1986, ZÜRN 1997a).

Short-term (up to one day) tilt and strain is dominated by tidal deformations and is at the order of 10^{-8} to 10^{-7}, which correspond to inclinations of 0.002" to 0.02" and lengths changes of 0.01 to 0.1 μm/m. *Long-term* effects of tectonic origin generally are only at the order of a few 10^{-7}/year. Episodic effects related to seismic or volcanic events may reach the same order of magnitude and more over a few hours to a few weeks and months. Consequently, the instrumental *sensitivity* of tilt and strainmeters should be at least about 10^{-9} to 10^{-10}, and the stability with time should be better than 10^{-7}/year.

Tiltmeters measure the inclination of the earth's surface with respect to the local vertical. Two mutually perpendicular sensors are needed in order to completely determine the tilt, which are usually orientated in the NS and EW-directions. Tiltmeters have been designed as horizontal and vertical pendulums, bubble levels, and long water tubes (ZÜRN et al. 1986).

Horizontal pendulums consist of two nearly vertical threads that support an approximately horizontal beam with an attached mass (*Zöllner* suspension), Fig. 5.65. Because of the small inclination of the rotational axis with respect to the vertical, a tilt of the support (basis about 30 cm) or a plumb line variation cause a strongly amplified angular deflection (astatization), which may be further enlarged optically. Fused quartz (*Verbaandert-Melchior* pendulum) ore metallic alloys are used as pendulum material in order to keep thermal effects small. Calibration is performed by controlled tilting of the instrument (VAN RUYMBEKE 1976). Among the *vertical pendulums* is the Askania borehole instrument, which may operate in depths of 20 to 60 m. The pendulum's (length 60 cm) suspension allows it to swing freely, and the deflections are sensed by two three-plate capacitive transducers installed at right angles to each other. The pendulum is calibrated by displacing a small mass over a known distance (FLACH 1976). Short-base *bubble levels* with electric sensors and *water-tube tiltmeters* with lengths of several 100 m have also been used. For the water-tube tiltmeters, water level variations at the endpoints of the tube are measured by capacitive or interferometric methods (KÄÄRIÄINEN 1979, SLATER 1983).
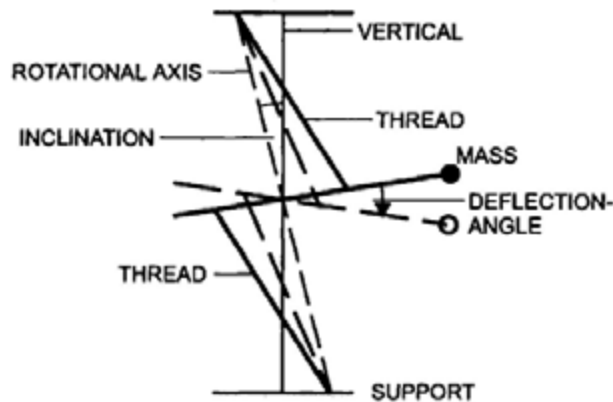
Fig. 5.65. Zöllner horizontal pendulum principle

Strainmeters measure relative displacements of the earth's crust (KING and BILHAM 1973). For a complete determination of the strain tensor, which contains six independent components, a strain meter array should be arranged with orientation in different spatial directions, although in most cases only horizontal strainmeters have been installed.

Short-baseline strainmeters (baseline length 10 to 30 m and more) are represented by *invar wires* and fused *quartz tube* rods. One end is fixed to a rock, and the crustal displacement is measured at the other end by inductive or capacitive transducers (KING and BILHAM 1976). *Laser strainmeters* apply the Michelson interferometer principle. Operated in an evacuated tube, they can measure distances up to 1 km with nm resolution (BERGER and LEVINE 1974, WYATT et al. 1990).

Among the *instrumental errors* of tiltmeters and strainmeters are the uncertainties of the calibration (about 0.1 to 1 %) and the direct effects of temperature and air pressure variations, which are kept small by the selection of the material and appropriate shielding. Long-term drift effects are at the order of 10^{-6} to 10^{-7}/year and to a large part are due to problems inherent with the sensor-rock coupling. Effects induced by *meteorological* and *hydrological* variations (air temperature, air pressure, solar radiation, rainfall, groundwater) pose severe problems in interpreting the results, especially for tiltmeters. These disturbances have pronounced daily and seasonal periods but also happen at other time scales. Modeling of these effects is still in its infancy, but their influence can be reduced by installing the instruments below the earth's surface, in tunnels, mines, natural caves, and boreholes.

Geologic, topographic, and *cavity* effects may cause large local distortions of the tiltmeter and strainmeter data. This is due to variable rock properties including local fractures, irregular topography, and different cavity reactions to deformation. Local distortion may reach 10 to 15% and more, leading to non-

representative results (HARRISON 1976). Attempts to model these effects have been only partly successful (SATO and HARRISON 1990, WYATT et al. 1990, KOHL and LEVINE 1995). Consequently, in addition to carefully selecting the observation site, preference now is given either to short-base tiltmeters and strainmeters operating in deep boreholes or to long-baseline instruments. In the latter case, local effects are reduced by integrating over a large distance of some 10 to some 100 m.

Tiltmeter and strainmeter results primarily contribute to earth tide research in the short-periodic part and to the detection of anomalous tilt and strain related to coseismic and volcanic activity, cf. [8.3.3].