Zhiping Lu · Yunying Qu
Shubo Qiao

Geodesy

Introduction to Geodetic Datum and Geodetic Systems

Springer

Geodesy

Zhiping Lu • Yunying Qu • Shubo Qiao

Geodesy

Introduction to Geodetic Datum and Geodetic Systems

Springer

Zhiping Lu
Yunying Qu
Shubo Qiao
Department of Geodesy
Institute of Surveying and Mapping
Information Engineering University
Zhengzhou
Henan
People's Republic of China

ISBN 978-3-642-41244-8 ISBN 978-3-642-41245-5 (eBook)
DOI 10.1007/978-3-642-41245-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014939507

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Content Summary

This book systematically and comprehensively discusses and explains the fundamental issues in geomatics and in surveying and mapping, such as geodetic datums and geodetic control networks, geoid and height systems, reference ellipsoid and geodetic coordinate systems, Gauss and UTM conformal projections, plane coordinate systems, and the establishment of geodetic coordinate systems. It also deals with various relevant geodetic data collection techniques.

The book can be used as a general textbook for undergraduates majoring in geomatics and in surveying and mapping in higher education institutions. For the technicians who are engaged in geomatic and surveying engineering, this book is strongly recommended as a basic and useful reference guide.

Preface

Geodetic datums and geodetic systems play an important role in surveying and mapping engineering. Geodetic datums refer to the reference surfaces, reference points, and their relevant parameters in surveying and mapping, including coordinate datums, vertical datums, sounding datums, and gravity datums. They are the reference surfaces or points against which measurements are made and they provide the basis for establishing geodetic systems. Geodetic systems are the extension of different types of datums realized through establishment of the nationwide geodetic control networks, which include the geodetic coordinate system, plane coordinate system, height system, and gravimetric system.

Geodetic datums and geodetic systems, as the common foundation for every subject of geomatics and surveying and mapping, are regarded as the main topic of this book. The book is designed to be used either as a reference for teaching or for learning subjects related to geodesy, surveying engineering, or geomatics. Some specific parts are written to fill literature blanks in the related area. For instance, we have extended the terms of traditional formulae with computer algebra systems to meet the accuracy of modern geodesy and have described modern geodetic coordinate systems and so on. The framework and structure of this book are formed through decades of teaching practice. The contents are systematic and the chapters proceed in an orderly and gradual way.

In writing this book, the authors put effort into building a new textbook system, attempting to avoid piecing together bits of knowledge from different courses. Due to the rapid and continuous developments in the field, it was necessary to be selective and to give more weight to some topics than to others. The material selected is particularly well suited to university-level students in line with twenty-first century education and the training requirements for a basic knowledge of geodesy. Therefore, in this textbook particular importance has been given to the fundamentals and to applications. It is a textbook that integrates classical materials with modern developments in geodesy, and balances practical applications and pure theoretical treatments by additionally highlighting some important and cutting-edge research issues in the field. Therefore, students who intend to pursue further studies in the field of surveying engineering should also find it helpful.

The book consists of seven chapters, a bibliography, an index, and a list of abbreviations. Summaries of the individual chapters are listed below.

Chapter 1 provides an overview of the discipline's objectives, roles, classifications, history, and trends in the development of geodesy.

Chapter 2 introduces the methods and principles of geodetic data collection techniques such as terrestrial triangulateration, height measurement, space geodetic surveying, and physical geodetic surveying.

Chapter 3 discusses the concept of geodetic datums and the methods, principles, and plans for establishing horizontal and vertical control networks, satellite geodetic control networks, and gravity control networks.

Chapter 4 deals with the basic concepts of the theory of the Earth's gravity field, discusses the definition of height systems, and establishes the relationship of transformation between different height types.

Chapter 5 discusses the reference ellipsoid, its relevant mathematical properties, methods for reducing the elements of terrestrial triangulation and trilateration to a reference ellipsoid, and establishes the models to transform mutually between the geodetic coordinate system, geodetic polar coordinates, and geodetic Cartesian coordinate system.

Chapter 6 is devoted to the methods and models of Gauss conformal projection and the Universal Transverse Mercator (UTM) conformal projection and establishes the relationship between the geodetic coordinates on the ellipsoid and the coordinates on the projection plane as well as the methods for coordinate transformations. The projection of geodetic networks from the ellipsoid onto a plane is also discussed so that they can be computed in the projected plane coordinate system.

Chapter 7 considers the principles of establishing classical and modern geodetic coordinate systems, establishes the transformation models between different coordinate systems, and provides an overview of the geodetic coordinate systems in China and throughout the world.

This book has been revised and extended by Zhiping Lu and Yunying Qu based on the first edition of the book, which was published in the Chinese language in 2006. In writing and adapting the original Chinese edition, Zhiping Lu wrote Chaps. 1, 4–7; Shubo Qiao and Jianjun Zhang wrote Chaps. 2 and 3. The numerical examples and illustrations in the book were designed and constructed by Shubo Qiao, Zhiping Lu, and Yupu Wang. English teachers Yali Zhang, Wen Zhang, and Yanxia Li helped with parts of the translation of the manuscript. Ph.D. candidates Zhengsheng Chen and Lingyong Huang and graduate students Yupu Wang, Hao Lu, and Kai Xie helped sort out part of the manuscript, read the manuscript, and offered some suggestions for revision.

The three reviewers of this book are Prof. h.c. Dr. Guochang Xu of the German Research Center for Geosciences (GFZ), Potsdam; Dr. Timmen Ludger of the University of Hannover; and Prof. Dr. Jörg Reinking of Jade University, Oldenburg. Dr. Timmen Ludger also mailed and presented two books for our reference. A grammatical check and correction of English language has been performed by John Kirby from Springer, Heidelberg.

Upon completion of the book, I wish to acknowledge the help and encouragement from all individuals who have, in one way or another, been involved in its preparation and completion. Particular thanks are due to Prof. h.c. Dr. Guochang Xu of GFZ. During the author's time as a visiting senior scientist at the GFZ, Prof. Xu has provided thoughtful care and prudent academic guidance. He has also helped with proofreading the manuscript and organizing the reviews by Dr. Timmen Ludger and Prof. Jörg Reinking, whose reviews are invaluable. Without his assistance, such a book would never be available. Thanks are also due to Prof. Dr. Frank Flechtner and Dr. Christoph Foerster, the head and acting head of Section 1.2 of GFZ, for providing the author with suitable facilities such as a working room and computing and communicating devices during the author's 2–3 month high-ranking scientific visit to and cooperation with GFZ every year from 2010 to 2013.

I also wish to express my sincere gratitude to Prof. Jingnan Liu of Wuhan University (Academician of Chinese Academy of Engineering, CAE), Prof. Yuanxi Yang of Xi'an Research Institute of Surveying and Mapping (Academician of Chinese Academy of Sciences, CAS), and Prof. Qin Zhang of China Chang'an University. Their long-standing encouragement and support are highly appreciated and gratefully acknowledged.

My special thanks should also go to my colleagues and staff at Information Engineering University for their support and trust during my teaching, researching, and writing of this book. They are Prof. Weiqiang Zhang (President of Zhengzhou Institute of Surveying and Mapping), Prof. Xiaosen Zhang (Minister of the Training Department), Prof. Guangyun Li and Prof. Hongzhou Chai (Deans of the Department of Geodesy), Prof. Xiangping Ye (Dean of the Foreign Language Department), and Mr. Yanbin Guo (Director of the Office of Teaching Affairs). To all these individuals and organizations, I express my cordial thanks. Their consistent support and efforts are preconditions for the publication of such a book.

This book was partially funded by the National Natural Science Foundation of China (NSFC, No. 41274015 and No. 40874008) and the National High-tech R&D Program of China (863 Program, No. 2013AA122501).

Potsdam, Germany Zhiping Lu
August, 2013

Abbreviations

APSG	Asia-Pacific Space Geodynamics
APT	Asia-Pacific Telescope
ARP	Antenna reference point
AUSLIG	Australian Surveying and Land Information Group
BIH	Bureau International De L'Heure
BJS54	Beijing Coordinate System 1954
CAE	Chinese Academy of Engineering
CAS	Chinese Academy of Sciences
CCRS	Conventional Celestial Reference System
CDP	Crustal Dynamics Project
CEA	China Earthquake Administration
CGBN57	China Gravity Basic Network 1957
CGBN85	China Gravity Basic Network 1985
CGBN2000	China Gravity Basic Network 2000
CGCS2000	China Geodetic Coordinate System 2000
CHAMP	Challenging Mini-Satellite Payload
CIO	Conventional International Origin
CLEP	Chinese Lunar Exploration Program
CMONOC	Crustal Movement Observational Network of China
COGRS	Continuously Operating GPS Reference Stations
CORE	Continuous Observation of the Rotation of Earth
CORS	Continuously Operating Reference System
COSMOS	Continuously Operational Strain Monitoring System with GPS
CRL	Communications Research Laboratory
CSB	China Seismological Bureau
CTRF	Conventional Terrestrial Reference Frame
CTP	Conventional Terrestrial Pole
CTRS	Conventional Terrestrial Reference System
DMA	American Defense Mapping Agency
DORIS	Doppler Orbitography and Radio-positioning Integrated by Satellite
DOSE	Dynamics of Solid Earth

DSP	Double Star Exploration Program
EDM	Electromagnetic distance measurement
EOP	Earth Orientation Parameter
EOS	Electro Optic Systems
EPN	EUREF Permanent Network
ERP	Earth Rotation Parameters
ESA	European Space Agency
ESLW	Equatorial springs low water
EUREF	Regional Reference Frame Sub-Commission for Europe
EUROLAS	European Laser Consortium
EVN	European VLBI Network
FAA‘	Federal Aviation Administration
FGCS	Federal Geodetic Control Subcommittee
GEONET	GPS Earth Observation Network System
GFZ	German Research Center for Geosciences in Potsdam
GIS	Geographic Information System
GLONASS	Global Orbit Navigation Satellite System
GNSS	Global Navigation Satellite Systems
GOCE	Gravity Field and Steady-State Ocean Circulation Explorer
GPS	Global Positioning System
GPST	Global Positioning System Time
GRACE	Gravity Recovery and Climate Experiment
GRS75	Geodetic Reference System 1975
GRS80	Geodetic Reference System 1980
GSFC	Goddard Space Flight Center
IAG	International Association of Geodesy
IAGBN	International Absolute Gravity Base Station Network
IAU	International Astronomical Union
IDS	International DORIS System
IERS	International Earth Rotation and Reference Systems Service
IGS	International Global Navigation Satellite System Service
IGSN71	International Gravity Standardization Net 1971
IHB	International Hydrographic Bureau
ILRS	International Laser Ranging System
ILS	International Latitude Service
IPMS	International Polar Motion Service
ISA	International Service Agency
ISLW	Indian spring low water
ITRF	International Terrestrial Reference Frame
ITRS	International Terrestrial Reference System
IUGG	International Union of Geodesy and Geophysics
IVS	International VLBI Service for Geodesy and Astrometry
LAGEOS	Laser Geodynamics Satellite
LEO	Low Earth orbit

LLR	Lunar Laser Ranging
LLW	Lowest low water
LO	Local oscillator
LOD	Length of day
MLLW	Mean lower low water
MLLWS	Mean lower low water springs
MLW	Mean low water
MLWS	Mean low water springs
MOBLAS	Mobile Laser Ranging System
MSL	Mean sea level
NAD83	North American Datum 1983
NASA	National Aeronautics and Space Administration
NAVD88	North American Vertical Datum 1988
NCRIEO	North China Research Institute of Electro-Optics
NGS	National Geodetic Survey
NNR	No-net-rotation
NSFC	National Natural Science Foundation of China
OPUS	Online Position User Service
PPS	Precise Positioning Service
PRARE	Precise range and range-rate equipment
PRF	Pulse repetition frequency
PRN	Pseudo-random noise
RTK	Real time kinematic
SA	Satellite altimetry
SAPOS	German Satellite Positioning Service
SBSM	State Bureau of Surveying and Mapping of China
SELENE	Selenological and Engineering Explorer
SGG	Satellite gravity gradiometry
SI	International System of Units
SLR	Satellite laser ranging
SMBGSH	Surveying and Mapping Bureau of the General Staff Headquarters of the Chinese People's Liberation Army
SNR	Signal-to-noise ratio
SPS	Standard Positioning Service
SSC	Set of station coordinates
SST	Satellite-to-satellite tracking
SSL-hl	High-low satellite-to-satellite tracking
SSL-ll	Low-low satellite-to-satellite tracking
SZCORS	China Shenzhen CORS
TAI	International Atomic Time
TIGO	Transportable Integrated Geodetic Observatory
TLT	Theoretical lowest tide
TRF	Terrestrial reference frame
TRS	Terrestrial reference system

TT&C	Tracking, Telemetering and Command
UNCED	United Nations Conference on Environment and Development
USACE	Corps of Engineers of United States of America
USB TT&C	Unified S-Band Tracking Telemetering Control
USCG	United States Coast Guard
UTC	Coordinated Universal Time
UTM	Universal Transverse Mercator
VLBI	Very Long Baseline Interferometry
VSOP	VLBI Space Observatory Program
WAAS	Wide Area Augmentation System
WGS84	World Geodetic System 1984
WPLTN	West Pacific SLR Network

Contents

Author Profiles

Zhiping Lu graduated in geodesy from Zhengzhou Institute of Surveying and Mapping, China, and obtained his bachelor and master degrees in 1982 and 1990, respectively. He received his doctorate from Wuhan University in 2001. Having worked as a lecturer from 1982 to 1992, and as an associate professor from 1992 to 1997 at Zhengzhou Institute of Surveying and Mapping, he has been working as a professor (since 1997), Ph.D. supervisor (since 2000), and deputy director of the key laboratory of surveying and mapping and navigation (since 2011) at Information Engineering University, as adjunct professor at China University of Petroleum (2005–2008), and as a visiting senior scientist at the German Research Center for Geosciences (GFZ) for 2–3 months each year (since 2010).

Yunying Qu graduated from Northeast Normal University, China, in 2006 and obtained her bachelor and master degrees in 2006 and 2008, respectively. She has been working as a lecturer at Information Engineering University since 2008 and has dedicated herself to the field of surveying and mapping.

Shubo Qiao graduated in geodesy from Zhengzhou Institute of Surveying and Mapping, China, in 1996 and obtained his doctorate from Shanghai Astronomical Observatory of Chinese Academy of Sciences in 2008. He has been working as an associate professor at the Information Engineering University since 2009.

Chapter 1
Introduction

Geodesy is a subdiscipline of geomatics and surveying and mapping as well as geoscience. It plays a pivotal role in construction of the national economy geoscience research, and in the process of social informatization. Advancement of modern science and technology has allowed geodesy to undergo an epoch-making transformation, to break the temporal and spatial limitations of traditional classical geodesy, and to enter a new stage of development of modern geodesy, primarily of space geodesy.

This chapter briefly introduces the disciplinary objectives, applications, history, and trends in the development of geodesy.

1.1 Objectives and Classifications of Geodesy

1.1.1 Objectives of Geodesy

According to the classical definition given by F.R. Helmert in 1880, geodesy is the "science of the measurement and mapping of the Earth's surface." This definition has to this day retained its validity; it includes the determination of the Earth's external gravity field as well as the surface of the ocean floor. With this definition, which has to be extended to include temporal variations of the Earth and its gravity field, geodesy may be included in the geosciences and also in the engineering sciences (Torge and Müller 2012; Helmert 1880).

The objectives of geodesy, generated from and partially supplementing Helmert's definition, can be described comprehensively as, first, to determine accurately the positions of points on the Earth's surface and their variations and, second, to study the gravity field of the Earth, the shape and size of the Earth, and the geodynamic phenomena. The former is generally considered the practical objective of geodesy and the latter the scientific objective. These two objectives are closely correlated.

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_1,

To our knowledge, geodesy is both a foundational and an applied discipline. As an applied discipline, geodesy, in the subject catalogue of some countries like China, is a subbranch of geomatics and surveying and mapping. Geomatics and surveying and mapping mainly study all forms of the Earth's surface; therefore the shape and size of the Earth and its gravity field are studied and measured, and a unified coordinate system should be established to show the exact geometric positions of an arbitrary point on the Earth's surface. Hence, geodetic measurements usually need to be carried out before topographic mapping. On the other hand, as a foundational discipline, geodesy, in the subject catalogue of some countries like China, is a subbranch of geophysics. Geophysics is primarily concerned with the Earth's movement, state, components, acting force, and all kinds of physical processes. Hence, geodesy provides instant, dynamic, and quantitative spatial geometric and physical information with high accuracy and resolution, which serves as an important means of studying geodynamic phenomena such as the Earth's rotation, movement of the Earth's crust, and changes of sea surface, and is used for prediction of geological disasters.

1.1.2 Classifications of Geodesy

According to the scope of the geospace studied, geodesy can be classified into ellipsoidal geodesy (i.e., theoretical geodesy, higher surveying), geodetic control survey, marine geodesy, and engineering geodesy (i.e., plane surveying). Ellipsoidal geodesy studies the body of the Earth as a whole, determines the shape of the Earth and its external gravity field, and establishes the geodetic reference system. Geodetic control survey measures the coordinates and heights of a sufficient number of surface points within one or several countries in an appropriately chosen reference system and establishes a unified national geodetic network to meet the needs of topographic mapping and engineering construction. Marine geodesy establishes a geodetic control network on the Earth's surface covered by oceans to realize positioning on the sea surface and underwater and to measure the marine gravity field, sea surface topography, and marine geoid. Engineering geodesy determines the details on the Earth's surface regionally in a small area and usually refers to the horizontal plane for measurement. Ellipsoidal geodesy, geodetic control survey, marine geodesy, and engineering geodesy are closely related to one another. National geodetic control survey and marine geodesy need the geodetic constants and reference datums determined by global geodetic surveying in order to reduce the observational results taking into consideration the effect of the Earth's curvature and the gravity field. The results obtained from the national geodetic control survey and marine geodesy provide ellipsoidal geodesy with information of geometric and physical measurements of the Earth's surface. Engineering geodesy has to be connected with the national geodetic control network to bring its results into the national unified coordinate system.

According to the spatial–temporal attributes of the Earth that are being studied, geodesy can be classified into geometric geodesy, physical geodesy, dynamic geodesy, and integrated geodesy. Geometric geodesy adopts geometric methods to study the shape and size of the Earth. It projects the terrestrial geodetic control network onto the regular reference ellipsoid as the basis for calculating the geometric positions of surface points. Physical geodesy is concerned with the external gravity field of the Earth globally or regionally. It establishes the theory of the Earth's shape by physical methods and deals with the geoidal undulation relative to the Earth ellipsoid using measured data of gravity. Dynamic geodesy studies the regional and global movement of the Earth and makes physical interpretations by accurately measuring the time-varying positions of surface points and the gravity field of the Earth. Integrated geodesy combines geometric and physical space and deals with all geometric and physical observed quantities of geodesy in a uniform mathematical model within the spatial–temporal reference system.

According to the technical means of carrying out the fundamental tasks, geodesy can be divided into terrestrial geodesy (conventional geodesy, i.e., astro-geodesy), space geodesy (satellite geodesy), and inertial geodesy. Terrestrial geodesy uses optoelectronic instruments to carry out short-distance (usually shorter than 50 km) terrestrial geometric survey (triangulateration, leveling, and astronomic surveying) and gravity measurement to determine the horizontal positions and heights of surface points and calculate the local gravity field parameters indirectly. By observing extraterrestrial objects (artificial Earth satellites, quasar radio sources, etc.), space geodesy realizes the positioning of surface points, including the relative positioning and the absolute positioning, which is relative to the center of the Earth. It uses the satellite gravity technique to obtain information on the global gravity field. Inertial geodesy applies the principle of inertia of a moving object in mechanics to carry out the relative positioning of surface points and measure the gravity field parameters.

A modern geodetic technology system centered on space geodesy has already been formed that can provide more accurate and abundant geodetic data than the classical system. It has not only expanded the application area of geodesy in socio-economic development but also improved its status as a basic discipline in geoscience.

1.2 Applications of Geodesy

1.2.1 Applications of Geodesy in Topographic Mapping, Engineering Construction, and Transportation

The important functions of geodetic control network in topographic mapping are primarily:

1. To control error accumulation in mapping. Errors are inevitable in mapping, for instance, they arise when we depict a line of direction or measure a certain distance. They are hardly noticeable in small areas, but would gradually propagate and accumulate in mapping of large areas, greatly deviating the topographic positions and features on a map. If a geodetic network is used as the basis for mapping control, errors can be constrained between adjacent control points to avoid accumulation and propagation so as to ensure mapping accuracy.
2. To unify coordinate systems. National basic topographic maps are generally mapped subdivisionally by different departments at different stages in different places. Because the coordinate system of points in the geodetic control network is unified nationwide with homogeneous accuracy, missing or overlapping layers do not occur in mapping, ensuring a perfectly fine splice of adjacent map sheets to form an integrated map.
3. To resolve conflict between an ellipsoid surface and a plane. A map is flat, but the Earth is approximately like a rotating ellipsoid with a non-developable curved surface that would crumple or split if forced to flatten, indicating that one cannot directly map the features from an ellipsoid surface onto a plane. However, the positions of geodetic control points on an ellipsoid can be projected onto a plane via certain mathematical methods. Mapping on the plane can therefore be controlled according to these point positions on the plane.

In this case, geodetic control points with certain densification have to be established first for topographic mapping. Traditional geodetic surveying has lower efficiency, consumes more time, and requires a greater workforce and a huge investment. With the fast-paced development of the economy, the demands for various kinds of medium- and large-scale maps are increasing rapidly, meaning that quick and precise positioning and rapid mapping techniques are needed to provide a guaranteed product. Modern Global Navigation Satellite Systems (GNSS), such as a Global Positioning System (GPS), can locate the position of a point within 5–10 min (compared with several hours to days using traditional methods) with centimeter-level accuracy. GPS allows rapid large-scale mapping when used for aerial photography and surface auto-mapping systems.

In engineering construction, the important roles of geodesy are:

1. To build a mapping control network for large-scale topographic mapping at the project design stage. Designers design buildings and plan districts on large-scale topographic maps. Geodesy serves to establish the mapping control network as the basis for mapping control.
2. To build a construction control network during project construction. A construction survey is mainly used to set out the designed buildings on a map and make sure that they are built in the intended locations. With different projects, the concrete tasks of construction surveys differ. For example, the major task of a tunnel construction survey is to ensure that the tunnel dug from reciprocal directions runs through in accordance with the specified accuracy. During layout, the direction and distance of the installed instruments are both calculated

on the basis of the control network, and thus the construction control network has to be established beforehand with the required accuracy.

3. To build a special control network for observing deformation with the purpose of monitoring deformation of buildings during the operation stage after completion of the project. A change in the original state of the Earth's surface during project construction, together with the weight of buildings, would cause inhomogeneous changes in the ground base and its surrounding strata. Besides, the building itself and its foundation will also deform due to changes in the ground base. Such deformation, once exceeding a certain limit, would affect the normal use of the building or even jeopardize its safety. In some cities (such as Shanghai and Tianjin in China) overexploitation of underground water could cause land subsidence in downtown areas on a large scale and bring about damage. Therefore, during the operational phase after completion of the construction, deformation needs to be monitored for these buildings or downtown. In this case, a control network for deformation observation with high accuracy has to be established.

With regard to transportation, geodetic surveying and positioning technology has provided important guarantees for improving traffic efficiency and decreasing traffic accidents.

The requirements of transportation for quantity, category, quality, and real-time positioning information are dependent on the level of development of social production, economy, science, and technology. The navigation and positioning level of ancient transportation means is from several kilometers to tens of kilometers whereas that of today's air and ocean transportation is from several meters to tens of meters. Modern GPS equipment can provide real-time positioning with decimeter-level or even centimeter-level accuracy, which is highly significant for large airports with frequent take-offs and landings. Currently, the number of automobiles in the world is increasing rapidly. According to statistics, traffic accidents in recent years are mostly due to drivers' failure to quickly determine the positions of and distances between automobiles and to their lack of quick-response capability while passing obstacles. At present, automotive GPS auto-positioning display and response systems are applied extensively, which will effectively reduce traffic accidents caused by automobiles. Such installations are also needed for inland navigation in narrow channels and ports to avoid ship collision accidents. Satellite navigation and positioning ability with high efficiency and accuracy enables traffic accidents to be greatly reduced and transportation efficiency to be highly improved.

1.2.2 Applications of Geodesy in Space Technology

The launching, guidance, tracking, remote controlling, and return of spacecraft need two basic types of support from geodesy: one is a precise geodetic coordinate

system and accurate positions of surface points (e.g., the launch point and tracking station) in this system; the other is a precise global gravity field model and accurate gravity field parameters (gravity acceleration, the deflection of the vertical, etc.) of the surface points.

A geodetic coordinate system is used to describe a spacecraft's movement relative to the Earth, which is realized by a certain number of datum points with known precise geocentric coordinates distributed on the Earth's surface. Its establishment includes determination of the orientation of its coordinate axes and a normal Earth ellipsoid defined by four fundamental parameters (a, J_2, ω, and GM) (see Sect. 4.2.1). The space tracking, telemetering, and command (TT&C) network, composed of a TT&C station (including a TT&C ship) in the space project, is adopted to determine the moving state (orbit and attitude) and working state of the spacecraft. It controls and adjusts the moving state of the spacecraft, builds their normal state, and manages aircraft under their moving state over a long term. The precise position of the TT&C station in the geodetic coordinate system is accurately determined by geodetic methods while the position of the spacecraft is solved from the given station coordinates of the TT&C through measuring the radial distance, range rate, azimuthal angle, and the like between the TT&C station and the spacecraft during operation.

The gravity field model provides prior constraint on the gravity field for analysis, description, and design of all mechanical behavior of moving objects on the Earth's surface and in outer space. Precise satellite orbit determination relies on the level of accuracy of the known expansion coefficients of the disturbing gravity potential in its dynamic equation of orbital motion. The lower order Earth gravity field model can ensure the accuracy of the orbit determination of a low Earth orbit (LEO)-satellite at the decimeter level. Space microgravity, a marginal discipline, emerged with advancement of interplanetary exploration technology. It is concerned primarily with the microgravity effect of tested objects on the spacecraft and is based chiefly on the high precision Earth gravity field model.

1.2.3 Applications of Geodesy in Geoscience Research

The components, movement, and development of the Earth system are observed and revealed by different branches of geoscience from different aspects using different methods. Geodesy places special emphasis on the study of the Earth's geometric (spatial) characteristics and fundamental physical characteristics (the gravity field) and describes their changes. Plate tectonics was developed in the late 1950s and early 1960s and has led the revolutionary progress of geosciences, which is significant for establishing the scientific view of "mobilism" in geoscience. The progress of modern geodesy and the introduction of space geodesy are essential in fostering geoscience development because geodesy enables extensive acquisition of information about the Earth's movement and allows the fundamental status of geodesy to be strengthened more profoundly. Modern geodetic techniques have

been powerful in support of the research of mobilism and can provide more abundant and accurate information for current research in the geosciences. The contributions of modern geodesy are chiefly:

1. It provides not only precise geodetic information for the study of plate movement and crustal deformation but also provides new methods for establishing accurate kinematic models of recent plate movement and crustal deformation. Very long baseline interferometry (VLBI), satellite laser ranging (SLR), and GPS are able to measure the precise and relative velocity of plates with an approximate speed of 1 mm/year so as to calculate directly the Euler vector of the relative plate movement from actual data. In the past 20 years, a massive amount of data on plate movement has been obtained using geodetic techniques; the correctness of the modern plate movement model NUVEL-1 derived from geophysical and geological data has been tested and observational models have been established. At present, geodesy is determining global, regional, and local crustal movements with unprecedented space–time resolution, according to which the stress–strain model inside the plates can be established to test the truthfulness of the rigid plate hypothesis, to deduce the deformation inside the plates, and to provide the basis for the explanation of faulting, seismicity, and other tectonic processes. Some geology and structures cannot be explained currently by plate tectonics, which awaits further improvement. Geodesy will presumably make new contribution to this.
2. The variations in polar motion and the Earth's rotation velocity are linked to information about the Earth's structure and diverse geodynamic processes. The precision of the Earth's rotation parameters determined by space geodesy has been the most effective tool in extracting and differentiating such information. Based on certain models (circle structure hypothesis, elasticity and viscoelasticity hypothesis of the Earth's mantle and core, etc.) of the Earth's structure, the corresponding rotation equations can be established to study the precession, nutation, and polar motion of the Earth's three axes (axis of rotation, axis of figure, and axis of angular momentum); the model of the Earth's structure can be verified and modified by comparing the observed values and the values of theoretical inference. One example of this is the correction proposed by VLBI observed data to the nutation series IAU1980, which impelled restudy of the Earth model. Polar motion includes the free motion (Chandler wobble) determined by the Earth's elasticity that lasts for 410–440 days, the superimposed forced oscillation lasting for a year, the minor swing lasting nearly a day, and the low-amplitude swing that lasts for a long period of 25–30 years. The factors causing these wobbles in different periods are a central subject of modern geophysics and involve a series of very important issues such as the exchange of angular momentum between the solid Earth, the atmosphere, the sea, and Earth's core; tidal friction and dissipation; the change in rotation angular momentum caused by seasonal climate variations; the viscoelastic core–mantle structure; nuclear magnetic fluid dynamics (geomagnetic dynamo); and core–mantle electromagnetic coupling. The motivating factor for the length of day

(LOD) is considered to be approximately consistent with that for the wobble of polar motion. There are still many aspects and arguments about the aforementioned issues in geophysics. Modern geodesy has established and implemented many programs around the world for monitoring the Earth's rotation and accumulated a massive amount of observed data. Combined with more information on geophysics, meteorology, and oceanography, it is possible to gain a new understanding of the above-mentioned issues about the Earth's structure and dynamics or even make breakthroughs by precise analysis.

3. A more refined gravity field will be provided through a series of satellite gravity survey programs and a larger scale survey of land and marine gravity. This geodetic finding will provide important data for analyzing and understanding the Earth's structure and dynamics.
4. Applied space geodetic techniques (particularly satellite ocean altimetry) could monitor changes in the sea surface with high accuracy and determine the sea surface topography and its changes. Such information can be used to study meteorological and oceanographic issues like global warming, atmospheric and oceanic circulations, etc.

As a dynamic system, the Earth witnesses extremely complex dynamic processes. With its unique theoretical system and survey methods, geodesy provides quantitative and qualitative data concerning dynamic processes on all kinds of spatial and temporal scales and reveals the essence of dynamic processes in combination with other relevant disciplines of geosciences.

1.2.4 Applications of Geodesy in Resource Development, Environmental Monitoring, and Protection

Resource exploitation, especially energy development, is a pressing issue for today's rapid economic growth. Topographic maps of various scales and precise gravity data are indispensable basic data for the exploration of both land and marine resources. For instance, in the early 1980s, the Doppler satellite network established in Tsaidam Basin, northwest China, and the gravity survey there provided precise geodetic data for the exploration and development of the oil field. Geodesy is especially important for the exploration and development of undersea continental shelf oil–gas fields. The satellite radar altimetry data combined with gravity surveys by offshore ships and leveling between coastal tide stations can provide the marine geoid, sea surface topography, and a gravity anomaly map with high precision and resolution in offshore areas. The application of radio-positioning on the sea surface, particularly the GPS marine positioning together with the sonar subsea positioning, can establish the three-dimensional (3-D) marine geodetic control network and draw subsea large-scale topographic maps. The marine geodetic data coupled with marine geophysical data such as marine geomagnetic surveys and drilling rock specimen sampling enable estimation of the conformation and reserves of oil and

gas under the sea. The data can also provide a basis for accurate determination of oil well positions, maritime and underwater fieldwork, positioning (or restoration) of drilling platforms, distribution of subsea pipes, and installation or callback of underwater detectors. The real-time, rapid, and accurate characteristics of satellite positioning techniques can provide the necessary guarantee for dynamic information management, production, command, and decision-making, and safe and stable operation in resource exploration and exploitation. Geodesy runs through the whole process of resource development from exploration to exploitation. State-of-the-art geodetic techniques are crucially important for exploration and development of mineral resources, particularly energy resources from the oceans.

Global warming and marine and air pollution are global environmental concerns in today's world. Developing countries still have regional environmental deterioration problems, such as water-induced soil erosion, desertification caused by ecological imbalance as the forest cover shrinks and grasslands degrade, large amounts of dust in the Earth's atmosphere, frequent acid rain in industrial cities caused by the excessively high proportion of coal used as energy source, large areas of water pollution resulting from failure to control the discharge of industrial waste, and so on. Environmental deterioration not only endangers human living conditions and quality of life, but also seriously restricts economic development. The influence of global warming has attracted the attention of scientists worldwide. One case in point is that global warming has been listed as an important issue for discussion in the 1992 UN Conference on Environment and Development (UNCED). The greenhouse effect could make polar ice sheets melt, sea water density decrease, and sea levels rise on a global scale. This, if combined with the estimated global mean sea level rise rate of 3.1 ± 0.4 mm/year as observed by satellite altimetry, could cause environmental changes such as shoreline erosion, land decrease, and seawater-induced soil basification, which will, over time, greatly endanger the living conditions of coastal residents. Many coastal areas and islands might be inundated by seawater. The strategy is to take this seriously and to monitor precisely this process and control its human causes (e.g., reduce the amount of carbon dioxide emissions, prohibit deforestation, etc.). The most effective means of monitoring this global change is space geodetic surveying and the most important method is to use GNSS tide gauges and process the data with reference to the International Global Navigation Satellite System Service (IGS), so as to analyze sea level changes in the precise geodetic coordinate system according to long-term observational results. The recently implemented satellite gravity gradiometry programs, the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE), are able to monitor the gravity change caused by melting of glaciers and ice sheets. Another important tool for measuring the ice is CryoSat-2's specially-designed all-weather microwave radar altimeter, which is capable of detecting changes in ice thickness to within 1 cm. Such data can help predict the effect of the melting polar ice on ocean circulation models, sea levels, and the global climate.

Countries worldwide have realized that environmental protection countermeasures have to be taken along with development of the economy. Environmental issues are a global concern. The increasing shrinkage of the Amazon rainforest in

Brazil and the tropical rainforests in Southeast Asia, the African primary forest destruction, and the spread of desertification in some areas all adversely influence the global climate and could trigger inundation and drought disasters. Therefore, a global environment monitoring system needs to be established and each country should have a sound monitoring system. The main measures are to develop remote sensing satellites, to establish a dynamic geographic information system (GIS), and periodically to make accurate and quantitative assessments of environmental changes. Development of such a monitoring system needs the support of geodesy: the launch of near-Earth satellites requires precise Earth gravity field models, the launch and tracking stations need accurate geocentric coordinates, and the establishment of GIS requires information about positions of points and controls. Geodesy serves indirectly in this system but is still vitally important and indispensable.

1.2.5 Applications of Geodesy in Disaster Prevention, Resistance, and Mitigation

Natural disasters, especially earthquakes, floods, and severe tropical storms, usually bring huge damage and loss to human beings. According to the statistics of Ministry of Land and Resources of the People's Republic of China, the average financial losses caused merely by geological disasters in China have amounted to about 4300 million US dollars each year since 2008 and, in years with frequent disasters, the loss caused by all kinds of natural disasters can reach one-sixth of the Chinese national gross domestic product (GDP). Therefore, countries worldwide set great store on preventing and fighting disasters. At present, excluding tropical storms (which can by and large be forecast accurately), it is still hard to predict massive earthquakes successfully, which reflects the inadequacy of man's knowledge of the science of the Earth. There is still a long way to go to improve our ability to prevent and mitigate natural disasters, which is an important mission of the geosciences, including geodesy.

Modern geodetic techniques, especially space geodesy, will be increasingly crucial in research on the monitoring and forecasting of earthquakes. Most earthquakes are distributed along the plate subduction zones and active intra-plate fault zones. According to historical statistics of earthquakes, the seismic activity of a seismic zone has a certain statistical periodicity. Geological evidence on prehistoric earthquakes in the plate subduction zones of the northwest Pacific Ocean has already been recognized, and the results of geodetic studies on crustal strains are in accordance with the accumulation of elastic strain in the period between two earthquakes, which supports the view of the Earth's recurrence period whose physical ground is the elastic rebound theory. This theory also serves as the basis for using the geodetic method to monitor crustal strains in a seismic zone over a long period to provide information for the medium- and short-term forecast of

earthquakes. Geodesy can monitor the whole process of the preseismic, coseismic, and postseismic strain accumulation and release and it is possible to establish the mode of earthquake precursors, combining this with geophysical observational results from the borehole strain meter, station extensometer, creepmeter, etc. For example, the successful short-term prediction of the 1975 Haicheng (Liaoning Province, China) earthquake used obvious short-term earthquake precursors. The earthquake is related to global plate movements, so when the relative velocity obviously deviates from the mean velocity, indicating that the strain accumulation at the borders of plates is above average, an earthquake would probably occur. Some countries such as the USA and Japan have established dense geodetic deformation monitoring systems, including GPS, VLBI, and SLR stations on seismic zones; for example, the USA has distributed GPS automatic monitoring networks on the San Andreas Fault Zone. Of course, earthquake prediction is extremely complex. We are almost sure that there will never be a precise earthquake forecast. We might know that an earthquake is due to occur, but no information about date and time, position of the epicenter, depth of the hypocenter, or any other important data will be predictable within the next 100 years.

Geodesy is equally important in preventing other kinds of geological disasters, for instance, the monitoring of landslide and mudflow. In 1986, the Rockfall and Landslide Research Institute of Hubei Province, China Three Gorges University accurately predicted a destructive landslide near the Xintan Area of the Yangtze River through geodetic monitoring, successfully preventing casualties and greatly reducing the financial losses to residents.

El Niño can cause disastrous climatic changes of long duration. The anomalies of sea temperature distribution and ocean circulation cause an abnormal change in atmospheric mass distribution through the interaction between ocean and atmosphere, which has generated floods in some places and drought in others. Due to the exchange of angular momentum, the change in atmospheric mass distribution causes a change in the Earth's angular momentum and influences the Earth's rotational velocity. When the 1982–1983 El Niño occurred, the Earth's rotation slowed down. With the VLBI and SLR techniques used today, the changes in the Earth's rotational velocity can be accurately measured, which enables prediction of El Niño several years (say 3 years) ahead.

Disasters happen in the world every year—plane crashes, shipwrecks, traffic accidents, people missing in severe environments, and so on. How to conduct timely and effective rescue becomes people's primary concern. In the past, wireless SOS distress signals were used to seek help, but often the exact position of the site could not be determined so the speed of rescue would be affected. Now, a satellite rescue system has already been established internationally and uses GPS rapid positioning and satellite communication technology to allow international rescue organizations to locate the site rapidly and organize rescue activities in time.

1.3 Brief History and Trends in the Development of Geodesy

1.3.1 Brief History of Geodesy

Geodesy has been gradually formed and developed as man's knowledge of the Earth has increased (Vanicek and Krakiwsky 1986; Jiang 1992; Torge and Müller 2012).

Embryonic Stage

Before the seventeenth century, geodesy was still in its embryonic state. In the third century B.C., Eratosthenes from Alexandria first used the relationship between the length S and the corresponding central angle γ of arc AB of the circumference and the circle's radius R in geometry to estimate the length of the Earth's radius (see Fig. 1.1). Since the two endpoints A and B of arc AB are approximately on the same meridian, the survey later developed into meridional arc measurement on this basis. In 724 AD, under the guidance of Zhang Sui (monk Yixing), Nan Gongyue in China Tang Dynasty first measured a meridian arc of about 300 km in the present Henan Province. Other countries also did similar work one after another. However, due to the rough techniques and the primitive measuring tools, although presumably the most precise tools available in former times, the accuracy yielded was rather low. This can only be seen as man's initial attempt to measure the size of the Earth.

Formation of Geodesy

There was a great breakthrough in man's understanding of the shape of the Earth in the seventeenth century. After Isaac Newton's formulation of the law of universal gravitation in 1687, C. Huygens from The Netherlands, basing his findings on the law that the gravity value on the Earth's surface increases from the equator to the two poles, claimed in his *Discours de la cause de la pesanteur* in 1690 that the Earth is an oblate spheroid flattened at the poles. In 1743, A.C. Clairaut from France published *Théorie de la figure de la Terre* and proposed the Clairaut theorem, which adopted the gravimetric method to determine the shape of the Earth. The research of Huygens and Clairaut laid the theoretical foundation for study of the shape of the Earth from a physical point of view.

In addition, W. Snell from The Netherlands initiated triangulation at the beginning of the seventeenth century. This method enabled distance measurement between two points on the Earth's surface several hundreds of kilometers or even much farther apart, solving the problem of directly measuring the arc length on the Earth's surface. Later, invention of the telescope, micrometer, level gauge, and

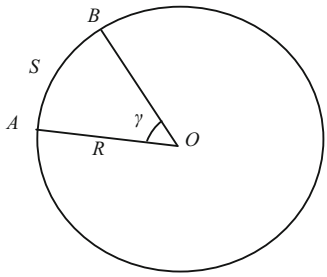


Fig. 1.1 Determination of the Earth's radius

other measuring instruments greatly improved the accuracy and laid technical foundations for the development of geodesy. So, we can claim that geodesy was established in the late seventeenth century.

Development of Arc Measurement

From 1683 to 1718, G.D. Cassini and J. Cassini from France measured the arc length of a $8°20'$ arc of the meridian across Paris by triangulation and deduced the semimajor axis and the flattening of the Earth ellipsoid from the lengths of its two arcs and the astronomical latitude measured at the two endpoints of each arc. Since the observation of the astronomical latitude did not reach the required accuracy and the two arcs were similar, they obtained a negative value for the Earth's flattening; namely, that the Earth is an ellipsoid elongated at the poles, which was the opposite of the deduction by Huygens based on the laws of mechanics. To clarify this doubt, in 1735 the French Academy of Sciences sent two survey teams to Lapland (situated on the border of Sweden and Finland) with a high latitude and Peru near the equator for meridional arc measurements. Their work finished in 1744 and the survey results from these two places certified that the higher the latitude, the longer the meridional arc per degree; namely, that the Earth is an ellipsoid flattened at the poles. Since that time, the physical assertion of the shape of the Earth is strongly supported by the results of arc measurement.

Another famous arc measurement was made by J.B.J. Delambre from 1792 to 1798, during which he measured the new meridional arc of $9°40'$ in France. From the data of this new arc and the meridional arc measured in Peru from 1735 to 1744, he deduced the length of a quadrant of the meridian and took one ten-millionth of it as the length unit named as one meter (1 m); this was the origin of metric system.

From the eighteenth century, to meet the needs of precise mapping, some other European countries started arc measurement and developed the layout method from along the meridional direction to the crisscrossed triangulation chains or triangulation networks. This work is no longer called arc measurement but astro-geodetic surveying.

In order to draw the *Complete Atlas of the Empire* (Huangyu Quantu), a large-scale astro-geodetic survey was conducted during Kangxi's reign in China's Qing Dynasty (1708–1718). The survey also proved that the meridional arc per degree at

higher latitudes is longer than that at lower latitudes. Kangxi also decided to determine the length of "Li" by setting the meridional arc length per degree as 200 Li.

Development of Geometric Geodesy

From the nineteenth century, many countries had started national astro-geodetic surveying, aimed at determining the size of the Earth ellipsoid, and, more importantly, providing accurate geometric positions of numerous surface points for national topographic mapping. To serve this purpose, a series of theoretical and technical problems had to be solved, which promoted the development of geometric geodesy. First, in order to test the abundant observational data from astro-geodetic surveying and eliminate their contradictions, on the basis of which to determine the most reliable result and assess observation accuracy, A.M. Legendre from France first published the theory of the least squares method in 1806. In fact, early in 1794, the German mathematician and geodesist C.F. Gauss had already used this theory to derive the asteroid orbit. Later, Gauss dealt with astro-geodetic survey results by the method of least squares and improved the method considerably, bringing about the adjustment of the observation method that is widely used in geodesy today and will be used in the future. Second, both the solution of a spheroidal triangle and the deduction of geodetic coordinates have to be done on the ellipsoid surface. Gauss proposed the solution to spheroidal triangles in his *Theorema Egregium* in 1828. Many scholars have worked out various formulae regarding deduction of geodetic coordinates. Gauss also published the conformal projection from the ellipsoid onto the plane in 1822 (which is why we call it the "Gauss conformal projection"). This is the best method to convert the geodetic coordinates into plane coordinates and is still extensively applied today. In addition, in order to use the results from astro-geodetic surveying to calculate the semimajor axis and the flattening of the Earth ellipsoid, F.R. Helmert from Germany proposed a method for solving the fittest ellipsoid parameters for the geoid of the survey area under the condition that the sum of squares of the vertical deflection of all astronomical points is the least in an astro-geodetic network. This method was later called the "area method."

Development of Physical Geodesy

Since Clairaut published *Théorie de la figure de la Terre* in 1743, the most important development in physical geodesy has been the Stokes theorem developed in 1849 by Sir. G.G. Stokes from the UK. According to this theorem, the shape of the geoid can be studied by terrestrial gravimetric results. However, the theorem first required reduction of the ground gravimetric results to the geoid, which was rather challenging. In spite of that, Stokes' theorem still promoted research on the geoid shape. About 100 years later in 1945, Mikhail Sergeevich Molodensky (*Михаил Сергеевич Молоденский*) from the former Soviet Union advanced

the Molodensky theorem, which enabled rigorous calculation of the distance between surface points and the reference ellipsoid (i.e., geodetic height) directly from terrestrial gravimetric results without any reduction. It is significant that the theorem allows direct and rigorous calculation of geodetic heights of the surface points by avoiding the geoid, which cannot be calculated rigorously in theory. Using geodetic heights, the observed values on the Earth's surface can be accurately reduced to the ellipsoid, which enables processing of astro-geodetic surveying data to avoid errors caused by inaccurate reduction. With the Molodensky theorem, the method of astro-gravimetric leveling and the normal height system emerged and were widely adopted by many countries.

Development of Satellite Geodesy

Up to the middle of the twentieth century, geometric geodesy and physical geodesy were well developed. However, measurements of the shape of the Earth and the Earth's gravity field were not satisfactory because astro-geodetic surveying can only be conducted on land, not across the oceans. There was only a limited amount of data from gravity surveys in areas like oceans, high mountains, and deserts. It was not until the launch of the first artificial satellite in 1957 that satellite geodesy began to emerge and geodesy developed into a brand new technique.

Not long after the emergence of artificial satellites, the satellite method was used to measure the flattening of the Earth ellipsoid precisely. More than a decade later, the measuring precision of the semimajor axis of the Earth ellipsoid reached ± 5 m, the coefficient of the spherical harmonics expansion of the Earth's gravity field was reliably deduced to the order of 36, and the global geodetic coordinate system had been established through satellite tracking stations. Modern GPS positioning techniques enable high-accuracy measurement of the geocentric coordinates of any surface point in the world geodetic coordinate system based on the precisely measured elements of satellite orbit. The use of the satellite radar altimetry technique to measure geoid undulation of the seawater has also yielded good results. The technique of satellite gravimetry is developing and also has great potential.

Development of Dynamic Geodesy

The Earth's crust is not still; it moves slightly and slowly due to lunisolar attraction, tectogenesis, and so on. Without precise surveying methods, such movement cannot be measured accurately. The VLBI technique was born in 1967. At the two ends of the baseline, which was thousands of kilometers long, radio-receiving antennae were established to receive synchronous signals from extragalactic quasar radio sources. By interferometry, the three baseline vector components of this baseline in the inertial coordinate system could be computed with centimeter-level accuracy. Quasar radio sources are extremely far away from the Earth and so have almost no angular motion relative to the Earth. Thus, from the given positions of some quasar

radio sources, it was possible to establish an extremely stable and therefore inertial space reference coordinate system. From long-term and repeated observations at short intervals, variations in the three baseline vector components could be computed and the polar motion, variations in the speed of the Earth's rotation, plate movement, and vertical crustal movement could be derived accordingly. Hence, the VLBI technique is an effective way to study the dynamic state of the Earth. In combination with the SLR technique and observations of the Earth's tide, dynamic geodesy emerged and provided strong support for geodynamics. With the maturation of GPS technology, since the 1980s GPS surveying has already become the chief method of dynamic geodesy.

1.3.2 Trends in the Development of Geodesy

Three hundred years have passed since the establishment of geodesy, and there have been considerable achievements in studies on the shape of the Earth, the Earth's gravity field, the measurement of surface point positions, and so on. The current trends in the development of geodesy are as outlined below. Exhaustive information about the trends of development is available in the literature (Schwarz 2000; Altamimi et al. 2011; Xu 2010; Xu 2012; Flechtner et al. 2010; Sideris 2009; Mertikas 2010; Drewes 2009; Pail et al. 2011; Plag and Pearlman 2009).

Modern Geodesy as Represented by Space Geodesy

The rapid development of modern technology, especially the development of laser, microelectronics, artificial satellites, extragalactic radio source interferometry, computers, and high-accuracy atomic frequency standard techniques, has given rise to an important breakthrough in geodesy and resulted in space geodesy with artificial satellites (signal) and extragalactic radio sources (signal) as the objects of observation. Such a breakthrough has enabled a relative accuracy of 10^{-6} to 10^{-9} of the distance and point position determination on a global arbitrary space scale and allowed determination of the 3-D position of a surface point with high efficiency within a few minutes or hours. This has radically broken the spatial–temporal limitation of classical geodesy. With the advent of quantum gravimeters and superconductor gravimeters, the ground gravity meter has also reached the microgal level (1 μGal $= 10^{-8}$ m/s^2), or even much higher accuracy. The satellite gravity technique involved in space geodesy can obtain gravity field information on a global scale, including the oceans. The breakthrough in technology has allowed the discipline to undergo an epoch-making revolutionary transformation and develop into a new stage of modern geodesy, as represented by space geodesy. The transformations are primarily:

1. Separated one-dimensional (vertical) and two-dimensional (horizontal) geodetic measurements have been developed into three- and four-dimensional measurements, which include time variables.
2. Measurement of the geometric and gravity field elements on the Earth's surface under the hypothesis of a static and rigid Earth has been developed into the observation of dynamic variations of the nonrigid (elastic and rheological) Earth.
3. Regional (relative) geodetic measurement in the local reference coordinate system has been developed into global (absolute) geodetic measurement in the unified geocentric coordinate system.
4. Measurement accuracy has been increased by two to three orders of magnitude.

These transformations have greatly expanded the research area of geodesy and have formed modern geodesy as opposed to classical geodesy.

Developing Towards Basic Research Areas in the Geosciences

The potential of development of modern geodetic techniques indicates the possibility of better monitoring of the Earth's movement and changes in its shape and potential field with sufficient accuracy on any spatial–temporal scale. The changes in the Earth's geometric and physical state are the result of its dynamic process under the effect of both internal and external forces. Geodesy serves not only to monitor and depict all kinds of geodynamic phenomena but also, more importantly, to explain their mechanism of occurrence and to predict their evolutionary processes. This is geodetic inversion, which includes the geophysical inversion of the crustal movement and changes in the Earth's rotation and gravity field. In other words, it means to infer, in reverse, the Earth's internal structural state, the source of forces, and the parameters of dynamic processes from the geodetic observational data of temporal variations. This intersectant research area between geodesy and its related geosciences has formed a new branch, dynamic geodesy. This is currently the most active interdisciplinary subfield of geodesy, whose development relies on the development of space geodesy and physical geodesy and is also closely related to the development of relevant geosciences. Dynamic geodesy is relatively independent, and its theoretical system and methods are still under establishment.

Modern geodesy is developing mainly towards and into the geosciences. Its basic tasks are:

1. To establish and maintain the inertial and terrestrial reference systems with high accuracy; to build and maintain regional and global 3-D geodetic networks, including the submarine geodetic network; to monitor in the long term the time-varying changes of these networks on a certain time scale; and to provide geodetic positioning and research of geodynamic phenomena with a highly precise terrestrial reference frame and a network of surface reference points.
2. To monitor and explain various kinds of geodynamic phenomena, including crustal movement, the Earth's rotational changes, the Earth's tide, sea surface topography, and the variations in sea level.

3. To measure the shape of the Earth and the fine structure of the Earth's external gravity field and its time-varying changes, and to interpret the observational results from a geophysical point of view.

These assignments will be accomplished along with development of relevant geosciences with the support of modern science and technology. Geodesy will be one of the frontier subjects in promoting the development of the geosciences.

Space Geodesy Will Dominate Future Developments in the Discipline

The dominant position of space geodesy in the future development of geodesy has already been determined by its wide application prospects and enormous potential. In terms of conventional mapping and general engineering control, GPS positioning techniques have basically replaced terrestrial measurement techniques using devices like theodolites and rangefinders. The former demonstrates superiority to the latter in accuracy, fieldwork efficiency, saving of workforce, and financial resources. In terms of the scientific objective of geodesy, the monitoring and research of various kinds of geodynamic and geophysical phenomena and processes will be the major assignments. These require geodetic techniques to realize this scientific objective on both the space and time scales; namely, to reach a sufficiently high spatial and temporal sampling rate. On the space scale, the objective is the ability to position on a regional and global scale more accurately and to determine the global gravity field with higher accuracy and resolution. On the time scale, the objective is the ability to monitor from crustal deformation by sudden earthquakes to long-term slow movements of plates; and, in populated seismic zones with intense tectonic movements, the ability to monitor successively and automatically. The monitoring accuracy for displacement is required to reach 10^{-8} to 10^{-9} (equivalent to ± 1 mm); the measurement of gravity anomaly is required to reach an accuracy of 1–3 mGal (1 mGal $= 10^{-5}$ m/s^2) with a resolution of less than 30 km. Concerning the current situation of science and technology, these requirements are only practicable with the vigorous development of space geodesy, primarily satellite geodesy.

The space geodetic techniques currently applied or developed mainly comprise the following: satellite positioning systems such as GPS, SLR, satellite altimetry, radio source VLBI, satellite gradiometry, and satellite–satellite tracking measurement.

Satellite Navigation and Positioning Techniques Have Expanded the Application Area of Geodesy

GPS techniques can provide static and dynamic objects with reasonably priced, highly efficient, continuous, and precise positioning and illustration of the state of motion. As an extensively applied technique, GPS has greatly extended the

application area of geodesy, except for application in the discipline of geodesy itself and the relevant geosciences. GPS positioning devices will become a necessity for people's socioeconomic activities and daily life in the information era.

Study of the Earth's Gravity Field Is Dedicated to Developing Satellite and Aerial Gravity Exploration Techniques and to Making a High Resolution Recovery of the Gravity Field of the Earth

In the last 30 years, the study of the Earth's gravity field has witnessed important progress such as the creation of satellite gravity technology and the emergence of absolute and relative gravimeters with microgal-level (1 μGal $= 10^{-8}$ m/s^2) accuracy.

In the entire framework of modern geodesy focusing on the study of basic geosciences, physical geodesy has closely integrated with space geodesy and constitutes the nucleus of geodesy, together dominating the development of the discipline. The level of refinement of the Earth's gravity field structure will be one of the major indicators of the future development of geodesy.

The development of gravity measurement will be dedicated to differentiating the shortwave spectrum of the gravity field and monitoring the temporal variations in the gravity field. The development of satellite gravity techniques will enable measurement of a global gravity field with an accuracy of 1–2 mGal and a resolution of 50 km. The accuracy of the latest fifth generation absolute gravimeter reaches ± 1 to ± 2 μGal; the accuracy of the superconducting (relative) gravimeter has reached 0.1 μGal. The accuracies of aerial gravity measurement and inertial gravity measurement are somewhere between ± 1 and ± 6 mGal, and are effective techniques for distinguishing the shortwave gravity field with a resolution of less than 50 km. Thanks to the development of gravimetry techniques, it has already become possible to monitor the temporal changes in the gravity field to provide new important data for the study of geodynamics.

Review and Study Questions

1. What are the technical and scientific objectives of geodesy?
2. What roles does geodesy play in topographic mapping?
3. Briefly summarize the history and trends of development of geodesy.
4. Analyze advantages of space geodesy compared with classical geodesy.

Chapter 2
Geodetic Data Collection Techniques

In order to accomplish the disciplinary tasks of geodesy, various geodetic data need to be collected extensively. This chapter briefly introduces the methods and principles of the data collection techniques that are commonly used in geodetic survey, such as terrestrial triangulateration, height measurement, space geodetic surveying, physical geodetic surveying, and so on.

2.1 Terrestrial Triangulateration

2.1.1 Angle Measurement

In establishing national geodetic control networks, it is often necessary to carry out horizontal and vertical angle measurements. The theodolite is an instrument for measuring angles with specific observational methods.

Horizontal and Vertical Angles

Horizontal Angle

In Fig. 2.1, A, P_1, and P_2 are three geodetic control points on the Earth's surface. Let A be the point of observation and P_1 and P_2 be the target points. Through point A, draw a plumb line AV (the direction of gravity) and a plane M, which is perpendicular to AV. The plane M is called the horizontal plane through point A.

The line of intersection Aq_1 between the horizontal plane M and the vertical plane containing the line of sight (line of collimation) formed by the plumb line AV and the line of sight AP_1 is called the projection of AP_1 on the horizontal plane, which is usually called the horizontal line (horizontal) of AP_1. Similarly, Aq_2 is the horizontal of AP_2. The angle β between Aq_1 and Aq_2 is known as the horizontal

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_2,

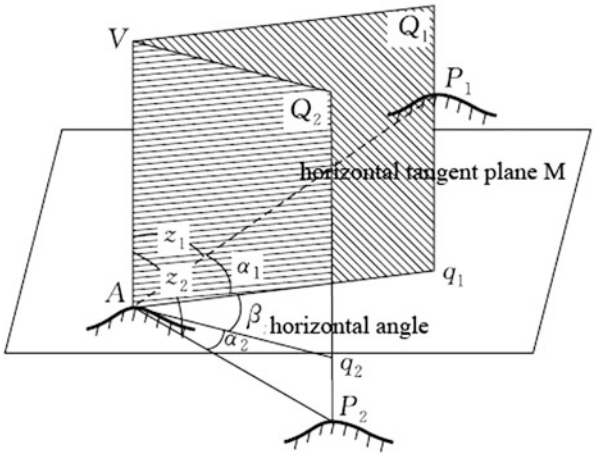


Fig. 2.1 Horizontal and vertical angles

angle to P_1 and P_2 from A. The horizontal angle is measured clockwise in the horizontal plane from 0° to 360°.

Vertical Angle

The angle between the line of sight AP_1 and its horizontal Aq_1 is called the vertical angle to the sighted point P_1 from A, denoted by α_1. Likewise, the angle between the line of sight AP_2 and its horizontal line Aq_2 is referred to as the vertical angle to the sighted point P_2 from A. Therefore, a vertical angle is the angle between the line of sight, which is the collimation axis of the telescope, and its corresponding horizontal, which includes the angle of elevation and angle of depression.

A vertical angle is measured in the vertical plane from 0° to ±90°, positive above the horizontal (see Fig. 2.1, α_1) and negative below (see Fig. 2.1, α_2).

The angles Z_1 and Z_2 between the plumb line AV and the lines of sight AP_1 and AP_2 are called the zenith distances from point A to the sighted points P_1 and P_2.

As illustrated in Fig. 2.1, the sum of the vertical angle and the zenith distance of a target is 90°, namely

$$\alpha + Z = 90^\circ. \tag{2.1}$$

According to this relation, the vertical angle and the zenith distance can easily be converted one to the other.

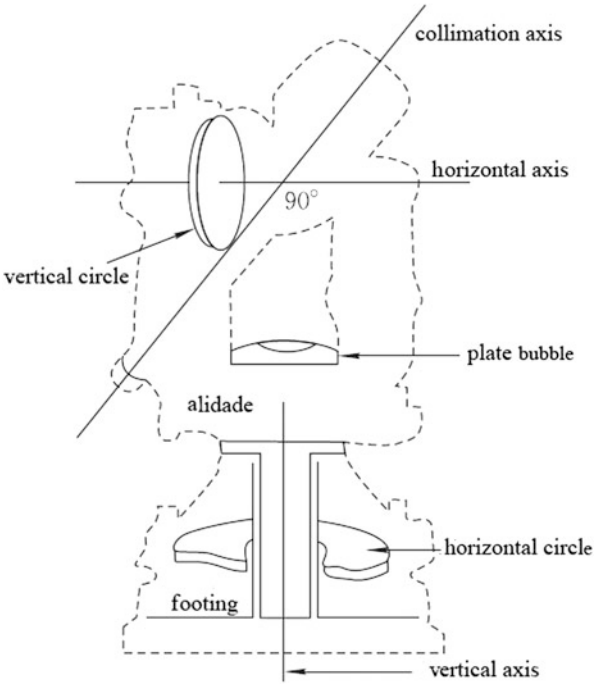


Fig. 2.2 Basic structure of a theodolite

The Theodolite

There are many types of theodolite, from the classical optical theodolite to the modern electronic theodolite. Theodolites are classified into J_{07}, J_1, J_2, and J_6 in China according to their precision. "J" is the initial letter of Chinese Pinyin (Jing Wei Yi) for theodolites and the subscript indicates their accuracy for angle measurement (mean square error). J_{07} and J_1 are of high accuracy and are used for the first- and second-order national control surveys, while J_2 is of medium-level accuracy for third- and fourth-order surveys (see, e.g., Xu et al. 1991).

A theodolite consists of the principal components shown in Fig. 2.2. These components are related as follows:

1. The vertical axis must be perpendicular to the axis of the plate bubble. When the plate bubble is centered, the vertical axis coincides with the plumb line.
2. The vertical axis must be at right angles to the horizontal circle and pass through its center. When the vertical axis is perpendicular, the horizontal circle is parallel to the horizontal plane through the point of observation and the angle measured in such cases will be the true horizontal angle.
3. The horizontal and vertical axes must be perpendicular and the collimation axis must be at right angles to the horizontal axis. So, when the vertical axis is perpendicular and the telescope is elevated or depressed, the plane formed by the collimation axis will be the vertical plane of sighting.

4. The horizontal axis must be perpendicular to the vertical circle and pass through its center. When the vertical axis is perpendicular and the horizontal axis is horizontal, the vertical circle is parallel to the vertical plane of sighting passing through the point of observation and the angle measured in such cases will be the true vertical angle.
5. Once the vertical circle index bubble is centered, the index for reading the vertical circle must be horizontal or vertical. Thus, the angle between the index of the reading scale and the collimation axis of the telescope will be the vertical angle.

The major parts of a theodolite are structured according to the relationships mentioned above. It is generally required that the relationship between three axes (the vertical and horizontal axes and the collimation axis) and between two circles (horizontal and vertical circles) be free of errors, which is crucially important.

Unlike the optical theodolite, the electronic theodolite provides a visual digital display of the circle readings instead of having to view through a reading eyepiece. It is therefore also called an electronic digital theodolite. The electronic theodolite is composed of optical mechanical devices, electronic sensor, microprocessor, etc. The configuration of its shafting, telescope, and clamp (tangent) screw are identical to that of the optical theodolite. The difference is that the electronic sensor is used to replace the index of the reading scale in the ordinary theodolite. Following the method of analog-to-digital conversion, it first receives electrical signals from the circle and then converts these electrical signals into angles and displays them on the monitor. Generally, there are kinematic and static angle measurements according to the rotation of the circle, and there are circular encoders and circular raster scales for angle measurements according to the different ways of circle graduating. Detailed principles are not discussed further here.

Methods for Observing Horizontal Angles

The Direction Method and the Closing the Horizon Method

For each set of observations, all directions at the station should be observed to get the angles. As shown in Fig. 2.3, the directions that need to be observed at station O are OA, OB, . . . and ON. Let OA be the starting direction (also referred to as zero direction). First, point A is sighted facing left and the reading is recorded. Then the alidade is rotated clockwise and the points observed in order from A to N and the readings noted, as one half of the full set of observations. Then the telescope is turned to face right, the alidade rotated anticlockwise, and the observations repeated facing right. The horizontal angles are then recorded again in reverse order from N to A as the other half of the full set. At this stage, one set of angles has been completed. Such a method is known as the direction method.

While using the direction method, before completing the direct-mode or reverse-mode readings, zero direction A will be measured again (called back-to-zero or being zeroed). Given the fact that each half of the whole set will combine and close up to the starting direction as a full circle, it is referred to as closing the horizon

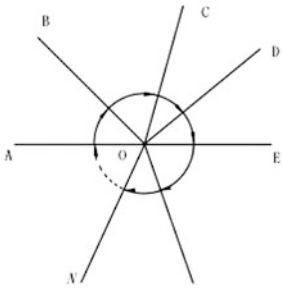


Fig. 2.3 The direction method

method. The purpose is for an additional check to see if any variations in the tribrach of the instrument have appeared during the direct or reverse mode of observations. The direction method and closing the horizon method are basically alike, and can be collectively called the direction method. When the number of the observed directions is equal to or less than three, the time consumed by one set of angle observations is quite short, and the direction method can be adopted (non-back-to-zero or not being zeroed); when the number of directions is greater than three (multiple angles), the closing the horizon method should be employed.

As to closing the horizon, whether the selection of zero direction is appropriate or not will exert some influence on both the accuracy and time of observation at the station. Therefore, the zero direction should be the one with appropriate length, good intervisibility, and clear target image.

Currently, the direction method is primarily used in angle measurement where lower precision is acceptable. Using the direction method, one obtains the values of all directions at a station. The value of the chosen zero direction is equal to zero. The angle between directions is the difference between the two directions.

Method of Angle Measurement in All Combinations (Schreiber's Method of Observation)

The direction method is simple and requires less effort in observation. However, the sides of the national higher-level control networks are of greater length, and the different target images cannot all maintain good quality at the same time. Moreover, the time consumed in one set of observations is rather long. It is therefore hard to achieve results with significantly greater accuracy. To overcome these deficiencies, the method of angle measurement in all combinations can be used. The major characteristic of this method is that it only measures the angle between two directions each time. In so doing, it is possible to overcome the difficulty in maintaining the clarity and stability of various target images simultaneously. Meanwhile, it also helps shorten the time used in one set of observations and makes it possible to achieve awesome results with higher accuracy, making it the preferred method for accurately measuring horizontal angles.

Each time, two directions are selected out of all the directions to be observed at the station and these are combined to form single angles; this is called the angle in

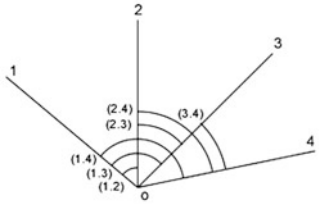


Fig. 2.4 Method of angle measurement in all combinations

all combinations, for example, if four directions need to be observed at the station, six single angles can be formed: (1.2), (1.3), (1.4), (2.3), (2.4) and (3.4) (cf. Fig. 2.4). If the number of directions at the station is n, then the number of angles in all combinations is given by:

$$K = \frac{1}{2}n(n-1). \tag{2.2}$$

For each set of observations, only a single angle is observed and the observation set for each combination angle is the same. The characteristic for such observation is that the alidade is rotated in the same direction for both direct-mode and reverse-mode reading. This is intended to better eliminate the errors due to backlash as the alidade rotates. However, for the entire observation and each set of observations for every single angle, the alidade should be rotated clockwise in a half set and counterclockwise in the other half so as to reduce other errors better. During each time period of observation this is achieved by either changing the rotational direction of the alidade when half of the set is completed or changing the rotational direction of the alidade in between the sets of observation alternately.

The above covers the direction method and angle measurement in all combinations. The key advantage of the direction method lies in its simple observation procedures and in the fact that less effort is required for operation. However, when there are multiple angles at the station, it will be hard to obtain clear target images in all directions. On top of that, the time consumed in a full set of observations is longer, which may make it more likely to be affected by external conditions, so difficulties may arise in achieving results with high accuracy. These are the disadvantages of the direction method. The method of angle measurement in all combinations has certain advantages, i.e., each single angle can be observed with a flexibly selected clear target for each set of observations, the time for observation is quite short, and the impact of external conditions on the observations is relatively small. However, the procedures for observation are relatively complicated; the combinations of single angles increase rapidly as the directions at the station increase and thus measurements require much more effort to perform. These are the drawbacks of such a method. Generally, the direction method is more applicable to angle observation with lower accuracy, whereas the method of angle measurement in all combinations is quite suitable for angle observation with higher accuracy.

The measurement of vertical angles requires little for observation and accuracy. What needs to be done is to observe the target using the crosshairs of the telescope during the direct-and-reverse modes of readings, which constitutes a complete set of observations. The vertical angle can be computed accordingly.

2.1.2 Distance Measurement

For hundreds of years, graduated tapes (measuring ropes, tape measures, and steel tapes) have been used to measure distance by means of direct comparison. The major flaw of such a method, however, is that it is easily subjected to the influence of topographic conditions along the survey lines. To obtain distance measurement with higher accuracy, one has to invest large amounts of human and material resources to choose and arrange the survey routes, which can be complicated and costly. Moreover, such distance measurement cannot be carried out if there are rivers, lakes, hillocks, or ravines along the survey lines.

With the progress of science and technology, in the 1940s a new type of distance measuring instrument came into being—the optical-electro distance measuring instrument, which was also the earliest type of electromagnetic distance measuring (EDM) device. Later, microwave, laser, and infrared EDM instruments emerged one after another. Even today, the total station electronic tacheometer that integrates angle and distance measurements is still available. It has created a new era of using EDM to replace the direct comparison method using graduated tapes and the indirect method using the optical tacheometer.

Principles of Electromagnetic Distance Measuring

In Fig. 2.5 an electromagnetic wave transmitted from a rangefinder placed at point A travels to the reflector at point B and back to point A, received by the rangefinder. Thus, the round-trip travel time t_{2D} of the electromagnetic wave between points A and B can be measured by the rangefinder. The distance (D) can be calculated according to the following formula:

$$D = \frac{1}{2} V t_{2D}, \tag{2.3}$$

where V is the velocity of propagation of the electromagnetic wave in the atmosphere, c is the velocity of the electromagnetic wave in vacuum, and n is the index of refraction (refractive index) for the electromagnetic wave:

$$V = \frac{c}{n}. \tag{2.4}$$

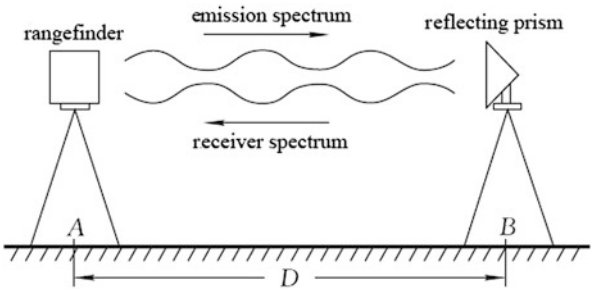


Fig. 2.5 Principles of EDM

The value of the index of refraction n is dependent on the wavelength λ of the electromagnetic wave and the meteorological elements of the atmosphere. The relationship between n and the air temperature t, the barometric pressure p, and the humidity e is expressed as (Duan 1996):

$$n = f(\lambda, t, p, e). \tag{2.5}$$

Knowing the wavelength of the electromagnetic wave and the temperature, pressure, and humidity of the atmosphere, the value of the refractive index n can be computed according to (2.5).

To sum up, the principle of EDM is to use instruments to measure directly or indirectly the round-trip travel time t_{2D} of the electromagnetic wave along the distance D and to measure the temperature t, the pressure p, and the humidity e at the same time to compute the distance according to the above formula.

It is clear that we can directly measure the distance between two points at the endpoint using the EDM method. Distances under any topographic conditions can be measured provided that the measurement range is reached and no obstacles interrupt the line of sight between the two points. Distances between mountains, rivers, and even planets for instance, using satellite laser rangefinder can also be measured directly, which can greatly reduce observation time.

Basic Methods of Electromagnetic Distance Measurement

There are three basic methods of EDM:

1. Method of Distance Measurement by Pulse (Pulse Method)

   The distance D to the target is attainable if we directly measure the travel time t between the transmitted pulse (dominant wave) and the reflected pulse (echo) from the target. With this, the distance can be obtained with a measurement performed only once, the measurement range varying from several kilometers to hundreds of thousands of kilometers. The precision generally reaches centimeter levels. Such a pulse method is chiefly used for measurements of low accuracy or

long distances, such as the front line tactical reconnaissance and distance measurements from Earth to the Moon and from Earth's surface to artificial satellites.

2. Method of Distance Measurement by Phase (Phase Difference Method)
   We can directly measure the phase difference between the transmitted signal and the echo to get the travel time. The measuring accuracy of such a method is better than the millimeter level and its measurement range is within dozens of kilometers. It is commonly used in precision distance measurement on the ground.
3. Method of Distance Measurement by Interference (Interferometric Method)
   This method adopts the physical principle of optical interference for precise distance measurement with higher accuracy than that of distance measurement by phase. Its precision generally reaches micrometer levels. It is normally used for calibrating distance measuring instruments and for precise short-distance measurement.

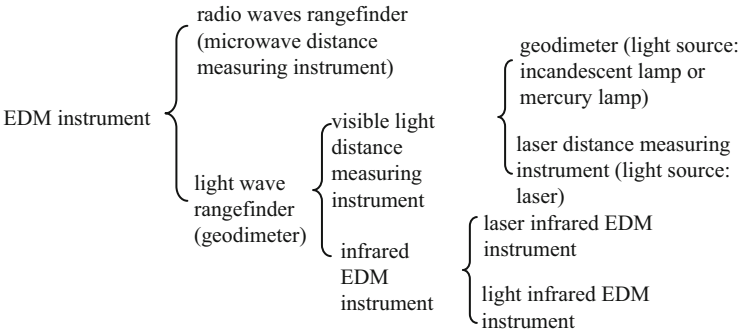
Classification of Electromagnetic Distance Measuring Instruments

EDM instruments can be classified into the following three categories according to the band of their carrier waves:

1. Microwave EDM Instrument
   The carrier wavelength ranges from 8 mm to10 cm in the microwave band.
2. Laser EDM Instrument
   The carrier is usually red visible light of 0.6328 μm wavelength and it is stimulated emission of radiation (i.e., laser emission).
3. Infrared EDM Instrument
   The carrier wavelength usually ranges from 0.75 to 0.95 μm and it is stimulated emission of radiation (i.e., laser emission) or spontaneous radiation (fluorescent light).

Generally, electromagnetic distance measurements involve distance measurements with radio waves and with light waves. Measuring distance using radio waves refers to microwave distance measurement. Light wave distance measurement includes two categories, one being visible light distance measurement and the other infrared distance measurement.

Visible light wave distance measuring devices can be categorized into two types. At the early stage of development, the instrument was based on spontaneous emission with an incandescent lamp or mercury lamp as its light source. In later models, a red laser ($\lambda = 0.6328$ μm) is stimulated by emission of radiation and the light source is generally a He-Ne-gas laser, as summarized below.



2.1.3 *Astronomical Measurement*

Astronomical observation is a technique utilized to determine the position of a point on the Earth's surface (astronomical longitude and astronomical latitude) and the astronomical azimuth by observing celestial bodies (especially stars). Astronomical observation is an ancient technology dating back to the era when human culture first took shape. To figure out direction and determine time and season, the sundial and gnomon were invented successively and Polaris was used for determining North.

Definition of Astronomical Coordinate System

In astronomical observation, the astronomical longitude, latitude, and azimuth of a surface point obtained with reference to its plumb line and the geoid are the representations of the surface point in the astronomical coordinate system.

As shown in Fig. 2.6, NS is the Earth's axis; the two points N and S, where the Earth's axis intersects the Earth, are the North Pole and South Pole, respectively; O is the geocenter; the plane $OWG'P'E$ through the geocenter perpendicular to the spin axis is the Earth's equatorial plane; P is a point on the Earth's surface; PK' is the plumb line direction of point P; the plane that contains the plumb line of point P is called the vertical plane of point P, in which the vertical plane $N'PP'S'K'$ parallel to the Earth's axis is called the astronomical meridian plane; $NGG'S$ is the initial astronomical meridian plane, also referred to as the astronomical meridian plane (see e.g., Xia and Huang 1995).

Astronomical Longitude

The astronomical longitude of a point P on the Earth's surface is the angle between the initial astronomical meridian plane and the local astronomical meridian plane of

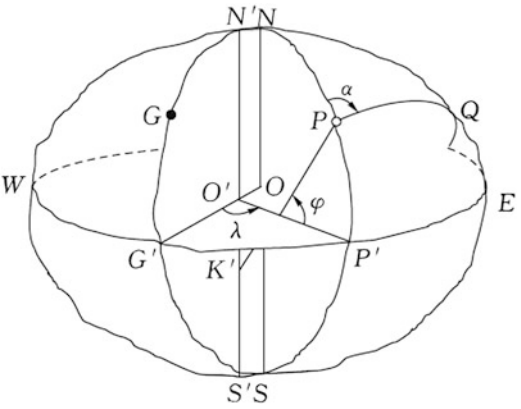


Fig. 2.6 Geoid and astronomical coordinate system

this point, denoted by λ. It is measured eastward or westward from the initial meridian and ranges from 0° to 180°; proceeding eastward is referred to as east longitude and westward is west longitude. East longitude is positive and west longitude is negative.

Astronomical Latitude

The astronomical latitude of a surface point P is the angle between the Earth's equatorial plane and the plumb line of this point, denoted by φ. It is measured northward or southward from the equator to the poles and ranges from 0° to 90°; proceeding northward is referred to as north latitude and southward is south latitude. North latitude is positive and south latitude is negative.

Astronomical Azimuth

Let P be the point of observation and Q the target point, then the vertical plane that contains point Q and the plumb line of point P is the plane of sighting of direction PQ. Therefore, the astronomical azimuth of direction PQ is the angle between the astronomical meridian plane of point P and the vertical plane of point Q, denoted by α. Its value is in the horizontal plane of the point of observation, measured from due north of the meridian clockwise from 0° to 360°.

Initial meridian is the intersection of the plane containing the meridian of 0° longitude with the Earth's surface. It is internationally agreed. In 1884, the International Meridian Conference decided to adopt the meridian passing through the Observatory at Greenwich, England (Airy Transit Circle) as the initial meridian, known also as the prime meridian or zero meridian.

Methods for Astronomical Observation

Traditional Methods for Astronomical Observation

Currently, astronomical observation usually adopts the traditional methods, mainly to determine time by receiving time signals emitted from the observatory and to record time by a chronometer. The instruments used in observations are primarily a Wild T4 theodolite and a 60° astrolabe, and the methods extensively applied are as follows:

1. The Wild T4 theodolite allows determination of the first-order astronomical latitude by applying Talcott's method.
2. The Wild T4 theodolite allows determination of the errors of a timepiece (time corrections) by the method of equal altitudes of two stars, one east and the other west of the meridian (Zinger's method), in order to determine the first-order astronomical longitude.
3. The hour angle of Polaris is applied for determining the astronomical azimuth.
4. A 60° astrolabe (composed of T3 and 60° prisms) allows simultaneous determination of the astronomical longitude and latitude of the second, third, and higher orders using the method of equal altitudes of multiple stars.

New Methods for Astronomical Measurement

The new methods primarily employ the GPS OEM (Original Equipment Manufacturer) board with time transfer service to receive satellite signals. Electronic theodolites are adopted for observation instead of optical theodolites; portable computers with advanced programming are used to replace the chronometer and timepiece for time comparison and timekeeping; and the autonomous recording and calculation of the observational data are also enabled. The currently adopted methods are as follows:

1. Use the method of observing multiple stars at approximately equal altitudes to determine the first- and second-order astronomical longitude and latitude simultaneously.
2. Carry out repeated observation using the method of hour angle of Polaris to determine the first- and second-order astronomical azimuth.

2.2 Height Measurement

2.2.1 Leveling

Principle of Leveling

Leveling is a method used for accurate determination of height difference between two points. The basic principle of leveling is that a precisely graduated staff is held vertically over the two points whose height difference is to be determined and then the scale readings are made with the horizontal line of sight. The difference between the two readings will be the height difference between the two points. As shown in Fig. 2.7, A and B are two surface points with unknown height difference. Leveling rods (leveling staffs) R_1 and R_2 are held vertically on each point while a level is placed at point S_1 in between these two points. From the horizontal line of sight, a reading of the rod R_1 is made as "a" (known as the backsight reading) and that of the rod R_2 is "b" (known as the foresight reading). Then the height difference h_{AB} between A and B is:

$$h_{AB} = a - b. \tag{2.6}$$

The height difference is positive when $a > b$ and negative when $a < b$.

Knowing the height H_A of point A, we can obtain the height of point B following $H_B = H_A + h_{AB}$. To determine the height H_P of an arbitrary point P, one needs to move the level to S_2 and the rod R_1 to point C after the height difference between A and B is determined. Then, the height difference h_{BC} between points B and C can be obtained. Likewise, the difference in height between A and P is $h_{AP} = h_{AB} + h_{BC} + \cdots$.

The height of point P is:

$$H_P = H_A + h_{AP}. \tag{2.7}$$

Such a method of transferring heights is referred to as geometric leveling.

Level and Leveling Rod

It can be seen from the principle of leveling that the leveling instrument should be developed to set up a horizontal line of sight. Therefore, the level should have a telescope capable of creating a line of sight (collimation axis) and a component that can direct the line of sight to the horizontal direction (a bubble is one of the simplest kinds). To make the line of sight horizontal and rotate horizontally, foot screws and a vertical axis are also necessary. Integrating these components as shown in Fig. 2.8 will constitute the simplest level. These principal components should satisfy the following conditions:

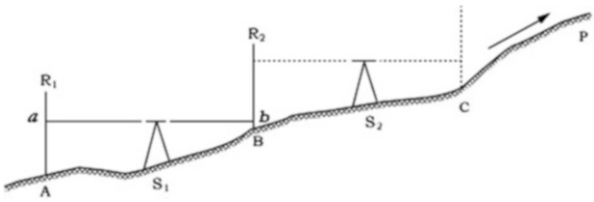


Fig. 2.7 Principle of leveling

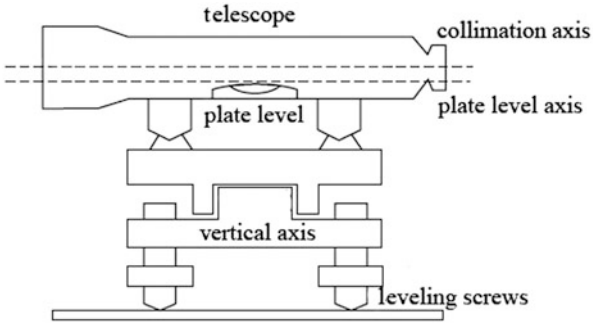


Fig. 2.8 Structure of the level

1. The line of sight is parallel to the axis of the bubble (level) tube.
2. The axis of the level tube is perpendicular to the vertical axis.

In this case, when the instrument is leveled with the bubble tube, the line of sight will be horizontal in all directions.

Precision dictates whether the leveling instrument can be classified as a precise leveling instrument or an ordinary leveling instrument. A precise leveling instrument is chiefly used in high-precision height measurement as in national first- and second-order leveling and precise engineering surveying. An ordinary leveling instrument is used in general engineering construction and topographic surveys. The major difference between these two types is that the precise leveling instrument has a built-in optical micrometer for accurate readings.

The leveling rod is an important leveling instrument that can be used to determine the difference in height between points.

A precise leveling rod has a 26 mm-wide and 1 mm-thick Invar strip placed in the grooves of the wooden part of the rod, with one end fixed to the base plate and the other to the metal frame at the top of the rod by a spring. Graduations of the rod are painted to fill the grooves cut in a scribed rule and the graduation lines are painted on the wooden part of the rod. The rod is approximately 3.1 m in total length.

The rod can be graduated at intervals of 10 mm or 5 mm, according to the measurement range of the level micrometer. Graduations are painted in two columns on the left and right sides of the rod.

The rod holder loop is also configured to both the back sides of the leveling rod to help hold it. The rod stand or stake is installed to keep the rod steady and upright.

Electronic Levels

The first electronic level was invented in March 1990 as a brand-new leveling instrument that integrated electronic technology, encoding, image processing, and computer technologies, and marked the direction for development of leveling instruments. Today, several corporations throughout the world are manufacturing electronic leveling instruments such as the DNA03 and DNA10 of Leica Microsystems, DiNi10 and DiNi20 of Carl Zeiss, and DL-101 and DL-102 of Topcon.

Different to the optical level, the rod face of the electronic level is graduated with a bar code and there also is an inbuilt digital image recognition and processing system. Using digital image processing technology, the image of the bar code can be processed and compared through a telescope, which enables the naked eye of the observer to be replaced by the array detectors (sensors). Observations (including clamping and reading) can therefore be completed automatically. In surveying operations, the leveling instrument needs to be only roughly leveled, the line of sight is automatically made horizontal by the compensator, the leveling rod is sighted, and the focus is adjusted. In such a case, by pressing the "measure" button, the reading of the rod and the distance between the rod and leveling instrument will be displayed on the monitor.

2.2.2 Trigonometric Leveling

Trigonometric leveling is a method for determining the difference in height between two ground control points by using the distance and vertical angle observed between the two points and then transferring the heights of ground control points. Compared with geometric leveling, trigonometric leveling is much simpler and more flexible. It is independent of terrain conditions and enables faster transfer of heights. The flaw in trigonometric leveling, however, is its slightly lower accuracy in determining heights. If controlled by leveling with sufficient density, trigonometric leveling can therefore not only ensure the accuracy of ground control point measurement but also overcome terrain constraints and improve work efficiency (see, e.g., Kong and Mei 2002).

Basic Principle of Trigonometric Leveling

As shown in Fig. 2.9, A and B are two points on the Earth's surface and their heights are H_1 and H_2, respectively. The vertical angle from A to B is α_{12}, S_0 is the horizontal distance between the two points, i_1 is the height of the instrument (HI), and a_2 is the height of the target (HT) of point B. The difference in height between A and B will be:

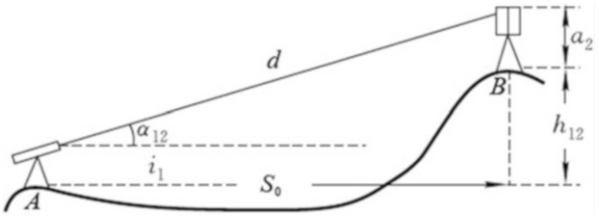


Fig. 2.9 Principle of trigonometric leveling

$$h_{12} = H_2 - H_1 = S_0 \tan \alpha_{12} + i_1 - a_2. \tag{2.8}$$

If the measured slope distance is d, the height difference is:

$$h_{12} = d \sin \alpha_{12} + i_1 - a_2, \tag{2.9}$$

which is the basic formula for computing the height difference using trigonometric leveling. Given the height of point A, that of point B becomes:

$$H_2 = H_1 + h_{12}. \tag{2.10}$$

EDM Height Traversing

Electromagnetic distance measurement (EDM) height traversing is also called precise trigonometric leveling. With the development of the electronic tachymeter, accuracies of angle and distance measurements have been greatly improved. The accuracy of distance measurement reaches over 1/100,000 and that of angle measurement can amount to 0. 5″, which provides favorable conditions for precise trigonometric leveling. At present, third- and fourth-order leveling can be completely replaced by EDM height traversing and, accordingly, in China specifications have been made by the departments concerned. Replacing leveling with EDM height traversing has proved to be notably economical in mountainous and hilly regions.

The methods of height traversing include reciprocal, leap-frog, and unidirectional. For the reciprocal method, the instrument is set up at each station to conduct reciprocal trigonometric leveling. The leap-frog method involves setting up an instrument midway between two targets. The targets remain at a particular change point. Observations are carried out in a pointwise manner. The targets should be set alternately and an even number of setups is used. This method is similar to leveling, except for using an oblique instead of a horizontal line of sight. The unidirectional method is based on the first and second methods, which is to observe twice with different heights of instrument at one station or to observe twice the two targets at each station. Tailor-made sighting vanes are used as the targets for all three methods described above.

2.3 Space Geodetic Surveying

2.3.1 GPS Surveying

Overview of GPS

Authorized to start in November 1973 by the US Department of Defense, GPS is a second-generation American satellite-based navigation system. It is accessible by the armed forces and cost 1,000 million dollars. It became fully operational in 1994 as the third greatest project after the Apollo lunar spacecraft and space shuttle. In surveying, navigation, guidance, precision positioning, dynamic surveying, time transferring, velocity measurement, and so on, GPS is convenient to use, easily operational in observation, precise in positioning, and beneficial economically. It has displayed its powerful functions and unparalleled superiority (Xu 2001).

The entire GPS consists of the space segment, the ground control segment, and the user segment.

Structure of GPS

The Space Segment

As shown in Fig. 2.10, the space segment is composed of 24 GPS operational satellites which form the GPS satellite constellation. Of these, 21 are navigation and positioning satellites and 3 are spares. These 24 satellites orbit around the Earth in six orbits at an inclination angle of 55°. Except at the two poles, theoretically more than four satellites can always be in view anywhere on the Earth's surface at any time. Orbiting at an altitude of about 20,000 km, each satellite makes one complete orbit every 12 sidereal hours. Every operational GPS satellite can transmit signals for navigation and positioning, which are then utilized by GPS users.

The Control Segment

The control segment is a monitoring system composed of several tracking stations around the globe. These tracking stations are categorized into master control stations, monitor stations, and up-link stations based on their functions. There is one master control station, which is located in Falcon Air Force Base in Colorado, America. Based on the observational data of GPS from every monitor station, it calculates the correction parameters of ephemeris and clocks of the satellites and uploads these data to the satellites through up-link stations. Meanwhile, it takes control of and gives instructions to the satellites. When one operational satellite goes wrong, it will dispatch a spare to replace the invalidated one. Additionally, the master control station also possesses the functions of monitor stations. There are five monitor stations. Apart from the master control station, the other four are

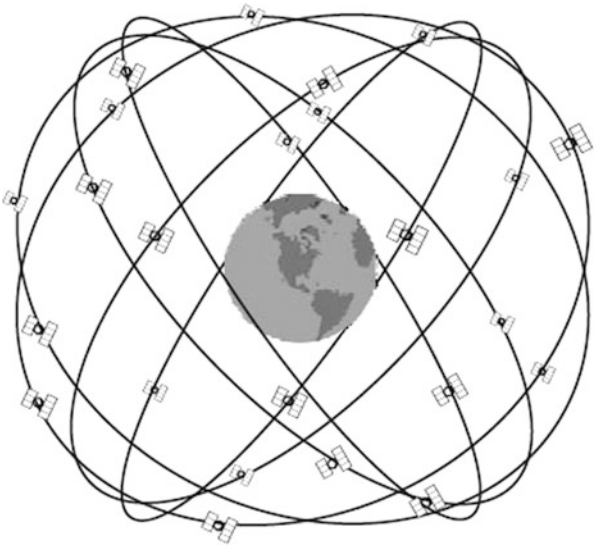


Fig. 2.10 GPS satellite constellation

located in Hawaii, Ascension Island, Diego Garcia, and Kwajalein. They are designed to receive signals from the satellites and monitor the satellite working status. The three up-link stations are located at Ascension Island, Diego Garcia, and Kwajalein. These stations upload to the satellites the correction parameters of ephemeris and clocks of the satellites computed by the master control station.

The User Segment

The user segment consists of the GPS receivers, the data processing software, and corresponding auxiliary equipment for users, etc. It is intended to receive signals sent by GPS satellites and to use these signals for navigation, positioning, and so on.

Signals From GPS Satellites

GPS satellites transmit carrier signals for civilian use at three frequencies: 1,575.42 MHz (L_1 carrier wave), 1,227.60 MHz (L_2 carrier wave), and 1,176.45 MHz (L_5 carrier wave). Their wavelengths are 19.03, 24.42, and 25.48 cm, respectively. Many signals, chiefly the C/A, P, and D codes, are modulated on carrier waves L_1, L_2, and L_5.

The C/A code, also known as coarse acquisition ranging code, is a pseudo-random noise code (PRN code) with a frequency of 1.023 MHz, The total code period contains 1,023 chips and lasts 1 ms. Different satellites can be distinguished by their PRN names because each satellite differs in its C/A code.

The P code, known as precision ranging code, is a PRN code at a frequency of 10.23 MHz.

The D code, known as the navigation message, at 50 bits per second, carries the position of satellites, status information, etc.

GPS Positioning Services

GPS offers two positioning services: the Precise Positioning Service (PPS) and the Standard Positioning Service (SPS).

PPS. Authorized users of the PPS, including the US military, certain US government agencies, and civil users specifically approved by the US government, need cryptographic equipment and special receivers. The positioning accuracy of PPS is several meters and the time accuracy reaches 40 ns.

SPS. For common civilian users, the US Government provides the SPS to take control of positioning accuracy. Users worldwide use the SPS without charge or restrictions. In the initial stages of GPS system implementation, SPS accuracy was intentionally degraded by the US Department of Defense by the use of so-called selective availability (SA). Under the effect of SA, the positioning accuracy of SPS is degraded to approximately 100 m and the time service accuracy is about 340 ns. In May 2000, the USA announced it was discontinuing the use of SA. At present, SPS provides a positioning accuracy of approximately 10 m and a time service accuracy of about 100 ns.

GPS Coordinate System and Time System

World Geodetic System 1984 (WGS84)

For a worldwide unified geodetic coordinate system, the US Defense Mapping Agency (DMA) has provided WGS60 since the 1960s and later developed the improved WGS66 and WGS72. WGS84, currently used by GPS, is a more accurate global geodetic coordinate system.

The coordinate origin of WGS84 is at the Earth's center of mass. Its Z-axis is the direction of the Conventional Terrestrial Pole (CTP), as defined by BIH1984.0. The X-axis points to the intersection of the zero meridian plane defined by the BIH1984.0 and the plane of the CTP's equator. The Y-axis constitutes a right-handed coordinate system.

The coordinates of GPS single point positioning and the baseline vector in relative positioning solution belong to the WGS84 geodetic coordinate system on which the GPS satellite ephemeris is based. However, practical measurement results often belong to a national or local coordinate system. In real applications, one needs to solve the transformation parameters in order to transform coordinates.

GPS Time System

GPS has a dedicated time system for precision navigation and positioning. The GPS time system, abbreviated as GPST, is provided by atomic clocks in GPS monitoring stations.

GPST belongs to the atomic time system. It has the same interval unit of one SI second as TAI (International Atomic Time), but a different point of origin from TAI, so there is an integer-second offset of 19 s in any instant of time between GPST and TAI, such that TAI − GPST=19 s. GPST was consistent with UTC (Coordinated Universal Time) at 0 h on January 6, 1980. Over time there is an offset of integral multiples of 1 s.

Features and Functions of GPS

With the employment of high-orbiting ranging systems, GPS defines its basic observed quantity as the distance between stations and the satellites. There are two main GPS measurement strategies for obtaining the observed quantity. One is the pseudo-range measurement, which measures the propagation time taken for the pseudo-random code to travel from the satellite to the user's receiver. The other is the carrier phase measurement, which records the phase difference between the carrier signals from the GPS satellites with Doppler frequency shift and the reference carrier signals produced by receivers. Pseudo-range measurement has the highest speed in positioning whereas carrier phase measurement has the highest precision in positioning. The three-dimensional position of a receiver can be deduced through the simultaneous pseudo-range or phase measurements of four or more satellites.

With the appearance of GPS, electronic navigation technology has entered a brilliant period. Compared with other navigation systems, GPS is mainly distinguished by the following (see Xu et al. 1998):

Continuous Global Coverage. Since there are enough GPS satellites in a reasonable distribution, at least four satellites can theoretically be observed continuously and synchronously from any point on the globe, which guarantees all-weather global continuous navigation and positioning in real-time.

Multifunction and High Precision. GPS can provide three-dimensional position, velocity, and time information continuously with high precision for all kinds of users.

High Speed in Real-Time Positioning. One-time positioning and velocity measurement of GPS receivers can be done within 1 s or even less, which is especially important for high dynamic users.

Remarkable Anti-interference Capacity and Adequate Confidentiality. Because of the employment of pseudo-noise spread spectrum technology in GPS, the signals from GPS satellites have remarkable anti-interference capacity and sufficient confidentiality.

GPS technology has developed into a new multidomain (land, marine, aerospace), multipurpose (in-transit navigation, precise positioning, precise timing, satellite orbit determination, disaster monitoring, resource survey, construction, municipal planning, marine development, traffic control, etc.), multimodel (geodesic type, timing-based, hand-held, integrated, vehicle-borne, ship-borne, airborne, satellite-borne, missile-borne, etc.), high-tech international industry. GPS has been widely applied in aerospace, fishing, guided tours, and agricultural production. It is said that "the applications of GPS are only restricted by human imagination."

GPS Measurement and Positioning Methods

GPS positioning methods are diverse. Users can use different positioning methods that are appropriate for their different purposes. GPS positioning methods can be classified according to different criteria as follows (see e.g., Liu et al. 1996):

According to Observed Values Adopted by Positioning

Pseudo-Range Positioning. The observed values adopted in the pseudo-range positioning are GPS pseudo-ranges, which can be a C/A code pseudo-range or a P code pseudo-range. This positioning method has the advantages of simple data processing, a low demand for positioning conditions, no integer ambiguity, and an easier realization of real-time location. The disadvantage is the low accuracy of observed values. The accuracy of the C/A code pseudo-range observations is generally 3 m and that of the P code pseudo-range observations is about 30 cm, which results in low accuracy of positioning results.

Carrier Phase Positioning. The observations adopted in the carrier phase positioning are GPS carrier phase observations; namely, L_1 carrier, L_2 carrier, or a linear combination of these. The advantage of carrier phase positioning is the high accuracy of observations, with a tolerance of better than 2 mm; however, it is complicated in data processing and has integer ambiguity.

According to Modes of Positioning

Absolute Positioning. Absolute positioning, also known as precise point positioning (PPP), is a positioning model in which one receiver is used. The absolute coordinates of the receiver antenna are determined in this mode. The mode is simple in operation, so it can be used in stand-alone operation. It is generally used for navigation and other applications of low accuracy.

Relative Positioning. Relative positioning, also known as differential positioning, employs more than two receivers to observe simultaneously in order to determine the mutual relationship between positions of the receiver antennae.

According to the Time Used for Obtaining the Results of Positioning

Real-Time Positioning. Real-time positioning, based on the observational data of the receiver, calculates the position of the receiver antenna in real time.

Non-Real-Time Positioning. Non-real-time positioning, also known as post-processed positioning, determines the position of the receiver antenna through the post-processing of the data received by the receiver.

According to the Receiver's State of Motion During Positioning

Kinematic Positioning. So-called kinematic positioning means that the position of the receiver antenna is changing with the time in GPS positioning. That is to say, in data processing, the position of the receiver antenna is seen to be variable over time.

Static Positioning. So-called static positioning is where the position of the receiver antenna in the whole process of observation remains the same. That is, in data processing, the position of the receiver antenna does not vary with the time. In a survey, static positioning is generally used for precise positioning. The specific observation model is to carry out static synchronous observations by multiple receivers at different stations for several minutes, hours, or days.

GPS Receiver

Navigation Receiver

GPS pseudo-range navigation is the most basic GPS service mode. GPS navigation uses the observation of distance (i.e., pseudo-range observations containing errors) to more than four satellites to determine the position of the receiver. Navigation-based GPS receivers (different from the phase measurement-based ones) generally carry out the pseudo-range and Doppler measurements only by C/A code or P code. They can receive navigation messages and calculate the position and velocity of the antenna in real time. Except for American military and authorized users, users can generally only use C/A code. As the most widely used receivers at present, they can be used in military and civil navigation, providing positioning with a medium degree of accuracy, and time transfer of relatively high precision.

Although GPS navigation receivers vary in type, their functions and operations are similar. The work process of a common navigation receiver is as follows:

1. Power on.
2. Wait for satellite searching. The receiver searches automatically for satellites that can be observed in the sky and locks onto the target, which will take a period of time varying from seconds to minutes according to the different types of receivers.

3. Display the positioning results. As soon as the receiver locks onto four (or more than four) satellites, it will begin the positioning and display the results. Position and velocity are generally displayed. The receiver constantly updates positioning and velocity results based on the selected data update rate.

Phase Measurement Receiver

Since the carrier wave is much shorter in wavelength than the pseudo-random code, in the case of the same resolution the observation accuracy of the carrier phase is much higher than that of the code phase. For example, the wavelength of the carrier L_1 is 19 cm, so the error of the corresponding distance observations is about 2 mm, whereas for the carrier L_2 the corresponding error is about 2.3 mm. The carrier phase measurement is the most accurate method nowadays and many companies have produced GPS phase measurement receivers of different types. The MacrometerV-1000 produced by the American company Litton Aero Service is a single frequency (L_1) phase measurement receiver and is the earliest manufactured for commercial use.

2.3.2 Satellite Laser Ranging

Satellite Laser Ranging (SLR), rising in the mid-1960s, is a new space geodetic technique that determines the distance between laser station and a satellite using the laser ranger to trace and observe the satellites installed with laser reflectors. To begin with, SLR employed the BE-C satellite. Then, in 1976 NASA launched a laser geodynamic satellite LAGEOS-1; in 1992 America cooperated with Italy and launched the satellite LAGEOS-2 to expand the observation range of SLR on the Earth. Meanwhile, France, the former Soviet Union, Japan, and Germany successively launched their SLR satellites. During its development over more than 40 years, the SLR system has improved from 1 m to the present 1 cm in distance accuracy. It is now one of the main technical means of precise satellite positioning as well as the most precise among the current various space observation technologies in terms of data sampling rate and absolute positioning. It not only plays a decisive role in establishing and maintaining the Global Geocentric Coordinate System (GGCS) but has also led to great achievements in the field surveying of modern plate motion, improvements in the Earth gravity model and geocentric gravitational constant, and the accurate measurement of Earth's rotation parameters (ERP), etc. (see e.g., Ye and Huang 2000).

Principles of Satellite Laser Ranging

SLR is a physical distance-measuring method, using the laser as its light source and the time of flight of the optical pulse for measurements. Its main features are:

1. The output power of the laser can reach orders of magnitude of 10^9 W and its optical energy density per unit area can be greater than that of the surface of the sun. Thus, the effective distance of the laser can reach the artificial Earth satellites tens of thousands of kilometers away, or even the lunar surface.
2. The laser spectrum is very sharp and has a halfwave width of about 5°, which benefits from adopting a narrow-band filter to eliminate sky background noise in a receiving optical system and to improve signal-to-noise ratio in observation.
3. The divergence angle of light beam output by the laser is very small, at about 1 mas. Through the optical system alignment, the divergence angle can be further compressed. Therefore, the light energy can still be concentrated within a very small scope far away.
4. The laser burst of a pulsed laser can reach a very small order of magnitude in width. Because the pulse width is one of the main factors in determining ranging accuracy, the laser ranging can be very accurate.

Due to the aforementioned characteristics of the laser, it is possible to realize long-distance laser ranging. There are three methods of laser ranging: the pulse method, phase method, and interference method. The pulse method is usually applied in SLR. Its basic principle is very simple: laser pulse signals are sent from a laser ranger placed at the observation station to a laser satellite equipped with a back-reflecting prism and go back to the receiving system of the rangefinder after being reflected by the tested satellite. If the time difference Δt between the sending and receiving of these laser pulse signals is measured, we can get the distance ρ between the satellite and the station according to the formula:

$$\rho = \frac{1}{2} c \Delta t, \tag{2.11}$$

where c is the velocity of light. Suppose the equation of motion of satellites in the Earth-Centered Inertial (ECI) Coordinate System is:

$$\dot{\mathbf{X}} = F(\mathbf{X}, P_d, t), \quad \mathbf{X}(t_0) = \mathbf{X_0}, \tag{2.12}$$

where $\mathbf{X}$ is the satellite's state vector at an instant of time t; $\mathbf{X} = (r, r_0)^T$ or $\mathbf{X} = \sigma$, σ being the six Keplerian elements; $\mathbf{X_0}$ is the satellite's state vector at initial time t_0; and P_d is the physical parameter to be estimated. Then the solution of (2.12) can be expressed as:

$$\mathbf{X} = Q\left(\mathbf{X_0}, P_d, {}^d t\right). \tag{2.13}$$

Suppose Θ_O is the satellite's observed quantity (i.e., the distance between the satellite and the Earth); its corresponding theoretical value Θ_C can then be defined as:

$$\Theta_C = \Theta(\mathbf{X}, \mathbf{R}), \quad \mathbf{R} = \mathbf{PNSR_0}. \tag{2.14}$$

where $\mathbf{R}$ and $\mathbf{R_0}$ represent the position vectors of the station expressed in the inertial system and in the Earth-fixed coordinate system, respectively, and $\mathbf{P}$, $\mathbf{N}$, and $\mathbf{S}$ stand for precession of the equinoxes matrix, the nutation matrix, and the rotation matrix of the Earth (see Vermeille 2002).

The above is the general principle of satellite dynamic geodesy. In practice, we should, based on different situations, purposes, and demands, select appropriate parameters as the adjusted quantities and keep the theoretical values of other parameters unchanged. Whichever dynamic geodesy is chosen, the satellite orbit is generally needed as the adjusted quantity, i.e., they all have a process of orbit determination and measurement.

SLR System

The SLR system consists of two main segments, a laser ranger on the ground and a laser satellite in space. The hardware devices of the ranger include seven parts: laser, telescope, electro-optical head, pulse position measurement system, time and frequency system, servo system, and computer; see Fig. 2.11 for an outline of the structure.

The working principle of the satellite laser ranging system is as follows:

The light pulse generated by the laser is led through the guiding optical path into the transmitter-telescope, which then emits the light beam to the target laser satellite after collimation. A small segment taken from the emitted light beam forms two electric pulses through the dominant wave sampling circuit. One is called the dominant wave pulse, which, as the enabling signal, is used to initiate the laser flight time interval counter. The other pulse is the electrical pulse, used to sample the clock and record the time of the laser emission. The laser pulse reflected by the satellite to the ground is received by the receiving telescope. Low-light detection equipment, which detects the reflected light, is set on the focus of the receiving telescope. The light is converted into electrical signals that are then amplified and reshaped into an echo pulse, which, as the stopping signal, stops the counter. In this way, the counter records the time interval $\Delta t = \tau - T$ between the dominant wave and the echo pulse (shown in Fig. 2.11), where Δt is the round-trip flight time of the laser between the station and the satellite.

The telescope of the ranger has three functions; namely, emitting, receiving the laser, and targeting satellites. We can design it as three independent telescopes or one telescope with three functions. Its time and frequency system has two functions: one is to provide a stable frequency source for counters, lasers, computers, and other devices (the stability of counters' frequency should be better than 10^{-10}); the other is to record the time for laser emission. The accuracy of time recording for

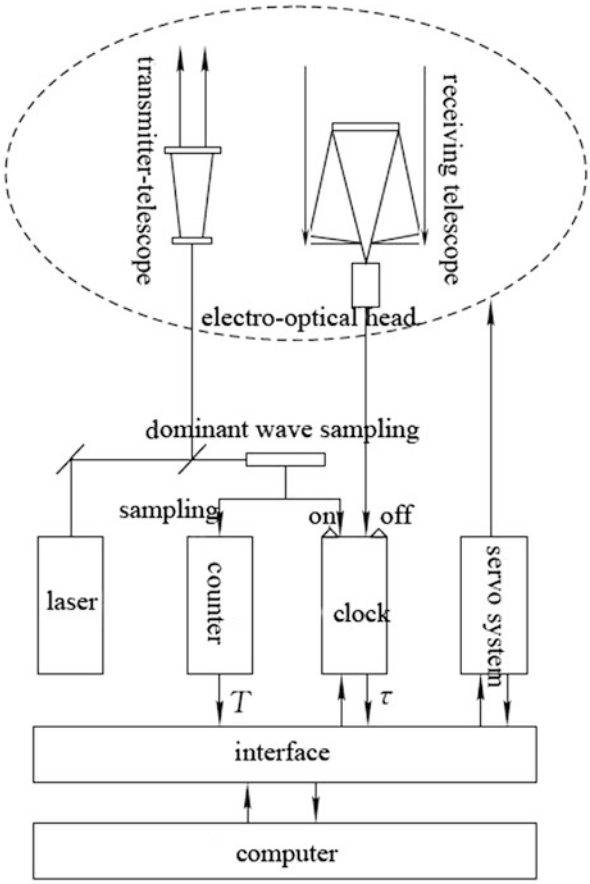


Fig. 2.11 Structure of the SLR system

the rangefinder with centimeter-level accuracy is 1 μs, and good quartz or rubidium clocks can meet the above requirements.

The laser ranger can only make an observation of satellites with dedicated reflectors. The laser reaching the satellite should return in the same direction as that of the laser emission. Such a reflector is also called a retroreflector, which is mainly composed of glass prisms. To achieve the accuracy required, the reflector must be carefully designed for the geometric shape and orbital altitude of dedicated satellites. To adjust the energy balance between the emitting laser and the receiving photons, the reflector should be big enough to reflect enough energy. In most cases, several single reflectors with 2–4 cm diameter set in a certain array can acquire the necessary energy. We should pay close attention to the alignment and adjustment of a single reflector lest the signal overlap causes a pulse distortion. The reflector is a passive device, easy to install on a satellite as an attachment. Therefore, many satellites are equipped with laser reflector arrays. Satellites LAGEOS-1 (in 1976) and LAGEOS-2 (in 1992) launched by NASA, and satellites Etalon-1 (in January

1989) and Etalon-2 (in May 1989) launched by Russia (the former Soviet Union), and others are dedicated laser ranging satellites for the applied research of geodynamics and geodesy. Satellites are especially suitable for geodetic research and observation because of their high and stable orbits, small area-to-mass ratio, symmetrical spherical shape, long accumulation time for observational data, etc. Figure 2.12 shows the shape of such satellites. They are spherical with a diameter of 60 cm, installed with 426 laser reflectors on their surfaces. They are still used for the common geodetic observation of SLR nowadays and have, among others, provided much data for establishment of the terrestrial reference frame and determination of the Earth rotation parameters. SLR has become one of the main techniques for satellite orbit determination because of its high precision distance measurement ability. Many reconnaissance satellites, meteorological satellites, Earth resources satellites, and oceanic satellites have all been equipped with laser reflectors so as to carry out more precise measurement and control of satellites by means of the SLR technique.

2.3.3 Very Long Baseline Interferometry

Very long baseline interferometry (VLBI) came into development in the late 1960s and is a radio interferometric observation technique that can combine two radio telescopes thousands of kilometers apart into a radio interferometry system with very high resolution. Since the line between the two telescopes is known as the baseline, VLBI is called very long baseline interferometry. Its resolution has now been improved to (the magnitude of) 0.1 mas with extension of the baseline. Because of the super-high resolution of VLBI, it has been widely applied in many fields like astronomy, geophysics, geodesy, and space technology for applications such as radio astronomy, accurate determination of the Earth's rotation parameters, crustal deformation detection, exploration of deep space and the ionosphere, etc.

Principles of Geodetic VLBI

Celestial bodies observed by VLBI are extragalactic radio sources, which are usually in deep space 100 million light years from Earth. When the electromagnetic wave radiated from the celestial bodies reaches the Earth's surface, its propagation distance is much further than that of the baseline in VLBI, so at this moment the movement of the wave front can be assumed to be parallel propagation and the wave is called a plane wave. On account of the different distances between the two antennae and a certain radio source, we get a distance difference L. Therefore, the time span of the radio signal from the same wave front to either antenna will be different, resulting as a time delay τ_g. According to the geometric relationship shown in Fig. 2.13 we get:

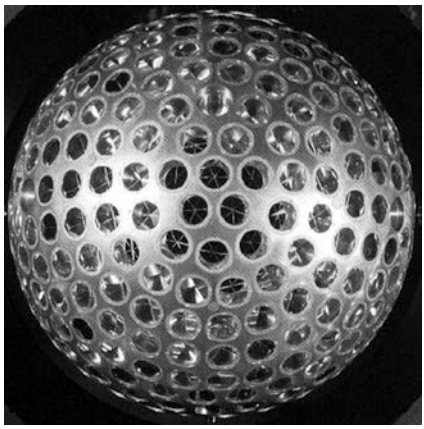


Fig. 2.12 Satellite LAGEOS. Source: NASA

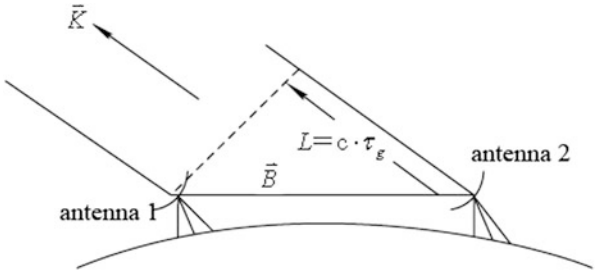


Fig. 2.13 Geometric principle of VLBI

$$L = c \cdot \tau_g, \tag{2.15}$$

where c is the velocity of light in vacuum. Suppose $\vec{B}$ is the baseline vector from antenna 1 to antenna 2 and $\vec{K}$ is the direction of the electrical source observed; then we get:

$$\tau_g = -\frac{1}{c}\left(\vec{B} \cdot \vec{K}\right). \tag{2.16}$$

Due to the movement of the Earth, the direction of vector $\vec{K}$ relative to the baseline vector $\vec{B}$ will change. Suppose τ_g is a time function, then we can denote its time derivative as the delay rate $\dot{\tau}_g$, namely:

$$\dot{\tau}_g = -\frac{1}{c}\frac{\partial}{\partial t}\left(\vec{B} \cdot \vec{K}\right). \tag{2.17}$$

The observed values of VLBI in geodesy are mainly the delay and the delay rate.

$\vec{B}$ and $\vec{K}$ in (2.16) and (2.17) must be expressed in the same coordinate system. However, the direction of the radio source is usually represented by the right ascension and the declination (α, δ) in the celestial coordinate system, and the baseline vector is defined as a vector $\vec{b} = (\Delta X, \Delta Y, \Delta Z)$ in the terrestrial coordinate system. As a result, in practical calculation, must be converted to a vector in the celestial coordinate system (see Dennis and Petit 2004):

$$\vec{B} = \mathbf{PNSW}\vec{b}, \tag{2.18}$$

where **P**, **N**, **S**, **W** stand for the rotation matrixes of precession, nutation, Earth's diurnal rotation, and Earth's polar motion, respectively.

For simplification, the influence of precession, nutation, and polar motion is not taken into account when discussing the principles of VLBI. Thus, (2.18) can be expressed as:

$$\vec{B} = \mathbf{R}_z(-\theta_g)\vec{b} = \begin{pmatrix} \Delta X\cos\theta_g - \Delta Y\sin\theta_g \\ \Delta X\sin\theta_g - \Delta Y\cos\theta_g \\ \Delta Z \end{pmatrix}. \tag{2.19}$$

Substitute (2.19) into (2.16) and (2.17) to obtain:

$$\tau = -\frac{1}{c}\left[\Delta X\cos\delta\cos\left(\theta_g - \alpha\right) - \Delta Y\cos\delta\sin\left(\theta_g - \alpha\right) + \Delta Z\sin\delta\right], \tag{2.20}$$

$$\dot{\tau} = -\frac{1}{c}\left[\Delta X\omega_g\cos\delta\sin\left(\theta_g - \alpha\right) + \Delta Y\cos\delta\cos\left(\theta_g - \alpha\right)\right], \tag{2.21}$$

where θ_g is Greenwich local sidereal time and ω_g represents the Earth's rotation speed.

The two equations above are the principle formulae for solving geodetic parameters by the use of the observed quantities of VLBI delay and delay rate. Through analysis of the formulae we know that the solution of VLBI parameters has the following characteristics:

1. VLBI delay and delay rate are pure geometric observed values that do not contain information about the Earth's gravitational field. Therefore, the acquisition of these observed values is not affected by the Earth's gravitational field.
2. VLBI, as a relative measurement technology, can only determine the relative position between two antennae, that is, the baseline vector; it cannot get the geocentric coordinates of each antenna. Thus, for determining the geocentric coordinates of the VLBI station we usually make observations through both VLBI and SLR in one station at the same time. Using the geocentric coordinates obtained by SLR as the datum, we can calculate the geocentric coordinates of other VLBI observations.

3. Because of its direct relation with the alteration of the Earth's rotation θg, the right ascension α of the radio source cannot be calculated independently from the observed quantities of the delay and the delay rate. As a result, VLBI alone cannot determine the origin of the right ascension of the radio source reference system and other technologies are required to do so.
4. The observed quantity of the delay rate does not contain the effect of the baseline component ΔZ, which cannot be calculated just by the observation of the delay rate. Besides, adding the data of delay rate to that of delay will not reduce the level of radio sources, which are to be observed to solve all the unknown parameters. In data processing and parameter solution, the delay rate is adopted only as a supplementary observation while the observed quantity of the delay is decisive.

The VLBI System

The VLBI system, as shown in Fig. 2.14, contains antennae, receivers, local oscillators, samplers, recording devices, related processors, and other units. The following is a brief introduction of the process of data collection of VLBI.

1. First, two antennae in the system receive the radio signal emitted by the observed radio source on the focus points of the antennae's paraboloid. Then, the feed source transfers the collected electromagnetic wave into high-frequency current and sends it to the receiver. The accuracy of VLBI observation (time delay, time delay rate) with regard to celestial and geodetic observation is directly proportional to the signal-to-noise ratio (SNR) of the system, and the SNR is directly proportional to the antenna aperture. Due to the weak signals from extragalactic radio sources, the VLBI antenna aperture is often over 20 m to obtain enough SNR.
2. The receiver's main function is to amplify this signal into a radio-frequency signal by using a high frequency amplifier and then convert it to an intermediate-frequency signal with a certain bandwidth through a mixer. In mixing, the mixer requires a local oscillating (LO) signal, which is provided by the station's local oscillator.
3. The intermediate-frequency (IF) signal from the receiver reaches the data recording terminal device, which nowadays employs an MK3 system or upgraded MK4 and MK5 systems. The MK3 recording system mainly contains two IF distributors, 14 video converters, a format cell data collection system, a magnetic tape recorder, and a computer that takes control of the data collection system and the recorder. The IF signal from the receiver is sent to the IF distributors and then to these 14 video converters, which convert the signals in different IF frequency ranges into video signals (also named Base Band) at 0–2 MHz that can be recorded by the magnetic tape recorder. The video signal output by the video converters is sent to the format cell. The main function of the format cell is to digitalize the 0–2 MHz video signals through one-bit sampling; then a format encoder supplies the precise receiving time for each datum and

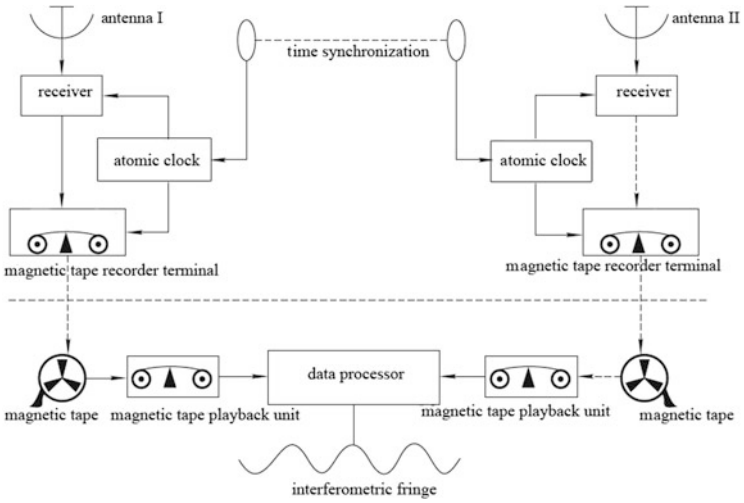


Fig. 2.14 VLBI system

encodes these data by rewriting signals and necessary information in a special format. Next, the formatted data are recorded on dedicated tapes in specialized format by the magnetic tape recorder. It should be noted that these 14 frequency converters have 14 independent LO signals, which will result in phase drift. So, phase calibration is required. A phase calibration system consists of a pulse generator, which transmit an impulse and inject it into the signals every microsecond. This pulse injection point is defined as the reference point of the delay.

4. Finally, the observed data recorded by the magnetic tape recorder are sent to the related processors. The processors playback the data and input them to the correlators of corresponding channels to carry out cross-correlation computations and acquire the related function value, i.e., the interferometric fringe. After that the computer uses its software system to obtain the required observed value of the time delay and the time delay rate by fringe fitting computation.

The Technique of Space VLBI

To improve the resolution of VLBI, the concept of space VLBI was proposed in 1970 and the establishment of the space VLBI system was also considered. By 1980, space VLBI became more mature in theory and technical realization. In 1997, the first space VLBI satellite (VSOP) in human history was successfully launched in Japan. Although space VLBI was proposed for astrophysics research, conceptually it has more advantages over ground-based VLBI for application in fields like geodesy. Therefore, it will become a more effective geodetic observation technology.

In light of the VLBI principle, there are no differences between space and ground-based VLBI. A space VLBI station can be seen as a component of a

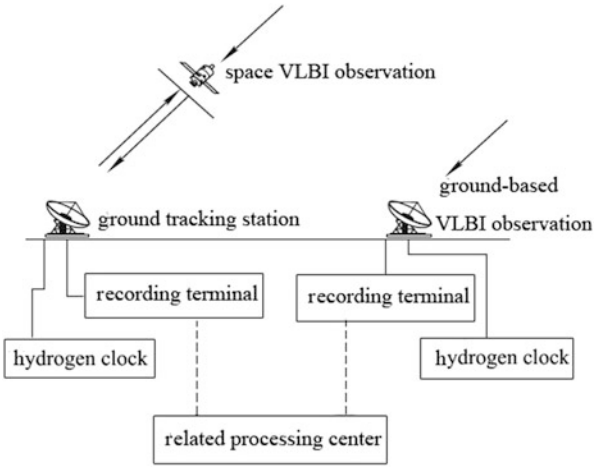


Fig. 2.15 Space VLBI system

ground-based VLBI net extending into space. It has the same function as a ground antenna, i.e., to receive signals from a radio source. Then, the required observational data for scientific research can be acquired through correlation processing of the signals received by both the space VLBI station and the ground antenna. However, space VBLI is different to ground-based VLBI in technical realization since the former places the antenna in space. The system of space VLBI is shown in Fig. 2.15.

Space VLBI is mainly distinguished by:

1. The phase of the space station's local oscillator is locked onto the hydrogen frequency standard of the ground tracking station. This frequency standard is sent to the space station from the tracking station through a radio channel (upwards).
2. The radio signal and other data received by the space station are sent back to the ground tracking station through a radio channel (downwards).
3. The space station must be equipped with highly precise systems for the attitude adjustment of antennae and for orbit control and detection.
4. The space station produces its own source of energy from received solar energy.
5. A ground support system of global coverage that can maintain uninterrupted communication with the space station is required.

The most significant technical advantage of the application of space VLBI to geodesy is to turn the geometric measurement of ground-based VLBI into dynamic measurement. It has been mentioned that the measurement completed by forming a baseline between two ground VLBI observations is a geometric one from the perspective of geodesy. Such measurement alone can only determine the relative position of the two stations but not their geocentric coordinates. Since the orbit of the space VLBI is described in a geocentric coordinate system and its movement is affected by various geodynamic factors, when adopting space VLBI a dynamic

measurement system can be formed by making a baseline between the space and ground stations so as to determine directly the geocentric coordinates of the ground-based station. Because all the VLBI antennae around the world take part in the space VLBI observation, a complete terrestrial reference system can be established independently using space VLBI technology itself. A VLBI station can not only be an observed object for various artificial satellite tracking stations on the ground, it can also be a space station in the orbit of artificial satellites to observe extragalactic radio sources directly. Thus, the direct connection and unification of the artificial satellite dynamic reference system and the radio source reference system can be realized. In addition, by means of space VLBI, an agreement can be made in VLBI about conversion between the Conventional Terrestrial Reference System (CTRS) and the Conventional Celestial Reference System (CCRS) to obtain a unified celestial and terrestrial reference system (i.e., a unified rotation and scale system with a commonly defined origin). Such unification of coordinate systems is of great significance for research in geodesy and other related fields.

2.3.4 Satellite Altimetry

In the 1980s, satellite altimetry (SA) appeared along with the application and development of computer technology, space technology, satellite telemetry, and remote sensing technology. SA employs microwave radar altimeters installed in satellites, radiometers, synthetic aperture radar, and other equipment to measure in real time the distance from a satellite to the ocean's surface, the effective wave height, and the backscattering coefficients, and to carry out research in geodesy, geophysics, and oceanography through data processing and analysis.

SA data can determine the marine geoid and solve the gravity anomaly of the ocean to compensate for the data gap in gravity measurement of marine areas. Therefore, SA plays an important role in establishing an Earth gravity field model with high accuracy and high resolution. The US Federal Geodetic Control Subcommittee (FGCS) noted that what the ocean altimetry satellite Seasat does within 3 months would take 200 years and cost 2 billion US dollars if done by marine gravimetry. Besides, SA data can also be used in oceanographic studies such as the measurement of the width, boundary, and velocity of ocean currents; tidal fluctuations; sea surface topography; and mean sea level changes.

The Basics

In SA, a microwave radar altimeter mounted on a satellite (the carrier) transmits microwave signals to the ocean's surface. This radar impulse reaches the ocean's surface and then returns to the radar altimeter by reflection. According to echo theory, we can obtain three observed quantities after the return of the radar pulse, including:

1. The round-trip time of the radar pulse going from the satellite to the ocean's surface and back to the satellite: the measured value of satellite altitude.
2. The waveform of an echo signal, which contains an ascending front, a flat top, and a decay area.
3. The amplitude of an echo signal, that is, the automatic gain control value of the signal. Based on the analysis of the waveform, structure, and round-trip time of the echo signal, we can obtain information like sea level altitude, sea level tilt, ocean currents, effective wave height, sea surface backscattering coefficients, and wind field (see Leuliette et al. 2004).

In satellite altimetry, the satellite is seen as a mobile platform on which a radar altimeter transmits a microwave pulse to the Earth's surface and receives the signal reflected back. Suppose that the altitude of the satellite is and the propagation velocity of the signal is c; we can then use Δt, the round-trip time of the radar signal observed, to calculate:

$$a = c\frac{\Delta t}{2}. \tag{2.22}$$

Because of water's good reflective properties, this method is particularly suitable for marine areas. The radiation of radar signals can instantaneously cover an annular region with a radius of thousands of kilometers on the sea surface (often referred to as the signal footprint). The size of the annular region is related to the spatial resolution of the incident microwave beam. So, the observed value is an elevation to the average instantaneous sea level. Its difference from the geoid height is $\overline{H}$. Assume that the satellite altitude with respect to the reference ellipsoid is h, which can be derived from the satellite orbit in the geocentric reference system through calculation. If neglecting additional corrections, we can simplify the basic satellite altimetry equation to:

$$h = N + \overline{H} + a. \tag{2.23}$$

Figure 2.16 shows that the radar altimeter can be adopted to scan the sea surface directly so as to scan the marine geoid approximately. Therefore, satellite altimetry is an effective method for directly drawing a geoid map. It is important that it can detect a very large marine area in a rather short time and make out a detailed sea level expression with a very high spatial–temporal resolution. $\overline{H}$ means a disturbance (noise) in establishing the geoid, but is an observed signal for the research of ocean dynamics. From the extensive analysis of $\overline{H}$, we can obtain an important understanding of the structure of the ocean floor and seabed features.

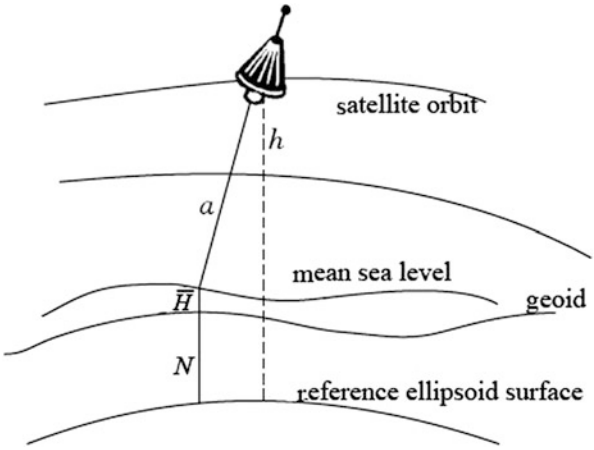


Fig. 2.16 Rationale of satellite altimetry

Satellite Altimeter and Its Operational Principle

A satellite altimeter is a satellite-borne microwave radar that usually consists of a transmitter, a receiver, a timing system, and a data collection system. It is generally 13.9 GHz in the emission frequency with 2 kW in transmitting power and works at an altitude of 800 km. Its radar antenna is parabolic with a diameter of 0.6–1 m. To guarantee at the same time the accuracy of measurement, the resolution, altitude, and other indices, the radar pulses transmitted must have a comparatively large time–frequency bandwidth. Thus, the satellite altimeter employs the pulse compression technique to transmit and receive pulses. The compressed pulses can be of nanoseconds (10^{-9} s) in width, which means that the pulse compression technique has solved a big problem in radio theory in that the width of pulses in the time domain and frequency domain cannot be enlarged simultaneously. The product of the width of pulses in the time and frequency domains is referred to as the compression ratio.

The operational principle of a satellite altimeter is as follows: The transmitter sends modulated compressed pulses to the Earth's surface with a certain pulse repetition frequency (PRF) through its antenna; after the reflection by the ocean's surface, the pulses return and are received by the receiver, which measures the time difference between the transmitted pulse and the received pulse. According to this time difference and the reflected waveform, we can determine the distance from the satellite to the ocean's surface. The width of the radar beam transmitted by the satellite is approximately 1°; as a result, when the radar beam reaches the ocean's surface, the radius of the signal track is about 3–5 km. Therefore, the distance measured by the altimeter is equal to the average distance from the satellite to the circular area with the radius of 3–5 km. On this basis, instrument adjustment, sea state correction, tropospheric refraction correction, ionospheric effect correction, and correction for periodic effects of the sea surface, etc. have to be taken into account.

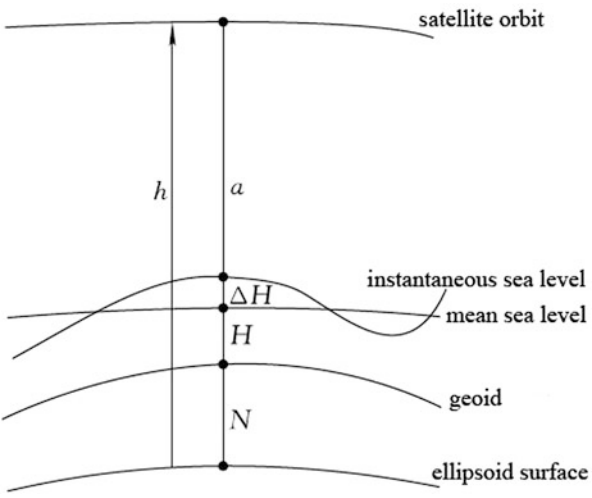


Fig. 2.17 Geometric relationship in satellite altimetry

The Observed Quantity in Altimetry and Error Analysis

Equation (2.23) is a simplified observation equation of the altimeter, which needs to be refined in practical application. The geometric relationship in SA shown in Fig. 2.17 yields:

$$h = N + H + \Delta H + a + d, \tag{2.24}$$

where h is the ellipsoidal height of the altimeter satellite based on orbit computations, is the geoid height, H is the sea surface topography, ΔH is caused by the instantaneous tidal effect, and a is the altimeter measurement.

$H + \Delta H = \overline{H}$ in (2.23). The quantity a is observed by the altimeter and requires atmospheric correction, which should be referred to the satellite's center of mass. The difference between the geoid and the mean sea level is called the sea surface topography, which can reach 1–2 m. The mean sea level is defined as a stationary sea surface not changing with time. The difference between the sea surface and the geoid is caused by differences in seawater salinity and temperature, a wide range of pressure differences, strong tidal currents, etc. For a resolution of better than 2 m, it is no longer valid to use mean sea level as a close approximation to the geoid. Besides, there are difficulties in connecting up the height systems obtained by a variety of tidal observational stations.

Errors and corrections of altimetric observations are categorized into three types; namely, the difference between the actual and the calculated orbit (the orbital error), the effects on signal propagation path, and the difference between the instantaneous sea surface and the geoid.

The orbital error mainly results from the accuracy of the Earth gravity model used in orbit calculation, errors in the tracking stations' coordinates, errors or limitations in the tracking systems, and errors in the model used for orbit

calculation. The most important impact generally comes from the Earth's gravity field. Since each satellite is only particularly sensitive to one certain subset of spherical harmonic coefficients, it is an effective method to develop a particular gravity model for the observed quantity of a particular satellite. For instance, use of the gravity model GEM10 in the altimetry satellite GEOS-3 improves the accuracy of the satellite's orbit from 10 m to 1–2 m. The tracking system is the second important factor to affect orbit accuracy. To acquire a high-precision orbit, orbit determination using on-board GPS is required, and the accuracy of the coordinates has reached several centimeters. Even so, the remaining orbit error is still much greater than the altimeter's accuracy. Consequently, we have to improve the orbit determination models, adopt some non-dynamic methods, etc.

The effects on the signal path can be categorized into instrumental errors and propagation errors. The major instrument effects include the distance between the phase center of the radar antenna and the satellite's center of mass, the propagation delay in the electronic circuit of the altimeter, and the timing error in the measurement system. In the manufacture of altimeters, these effects can be reduced to a minimum and can be estimated. All the effects of instrumental errors should be determined and kept under control when calibrating an altimeter in the test area with measured accuracy. The signal propagation error caused by ionospheric refraction is about 5–20 cm and depends on the ionization intensity, whose effects can be corrected by dual frequency. The effect of tropospheric refraction is about 2.3 m. Since only the observed quantity in the vertical direction is adopted, the effect can be well corrected by a proper atmospheric refraction model to reach an accuracy of several centimeters. Propagation errors also include the impact of actual sea conditions on the reflected signals.

The discrepancy between the instantaneous sea surface and the geoid can be divided into a time-invariant part H and a time-dependent part ΔH. Before determining the mean sea level using the altimetry observed quantity, the time-dependent component should be corrected. The wave-induced sea level change, which has been smoothed out in the altimeter's observation, can be negligible. Therefore, the correction contributor to be considered is mainly the sea level change induced by tides and changing atmospheric pressure fields.

The resolution and accuracy of the sea surface height measured directly from satellite altimetry can reach 5 km and 5 cm, respectively. However, under the effect of the sea surface topography, the tides, and errors in the environmental correction model, the accuracy of the sea geoid can rarely be better than ± 10 cm.

2.4 Gravimetry

2.4.1 Absolute Gravimetry

Absolute gravimetry (absolute gravity measurement) is a technique utilized to determine the gravity value (actually, gravitational acceleration) at a defined geometric point. There are two methods for absolute gravity measurement, one using a reversible pendulum, and the other by means of the free-fall motion of bodies. The second method of measurement has been the dominant method since the 1960s and will be discussed here.

Free-fall motion refers to the accelerated linear motion of a body along the plumb line under the action of gravity only. According to mechanics, if the gravity acceleration g in the course of motion along the plumb line is assumed constant (no gravity changes with height), then the equation of motion is:

$$l = l_0 + V_0 t + \frac{1}{2} g t^2, \tag{2.25}$$

where V_0 and l_0 denote the initial velocity of the falling body and the distance from the origin O, respectively, at the computational time $t = 0$, and l is the distance of the falling body from the origin O after a period of time t, cf. Fig. 2.18.

Two methods can be used to determine the gravity value by means of the free-fall motion of bodies: the free-fall method and the symmetrical rise-and-fall method (abbreviated as the rise-and-fall method). Their principles are discussed below.

Free-Fall Method

In (2.25), to avoid determining V_0 and l_0, it is necessary to measure from at least three positions. Assume at time t_1, t_2, and t_3 that the distances of the falling body from the point O are l_1, l_2, and l_3, respectively, as illustrated in Fig. 2.19, where the transverse axis indicates time and the ordinate axis indicates distance. With reference to (2.25), for each time period there will be a corresponding equation of motion which gives:

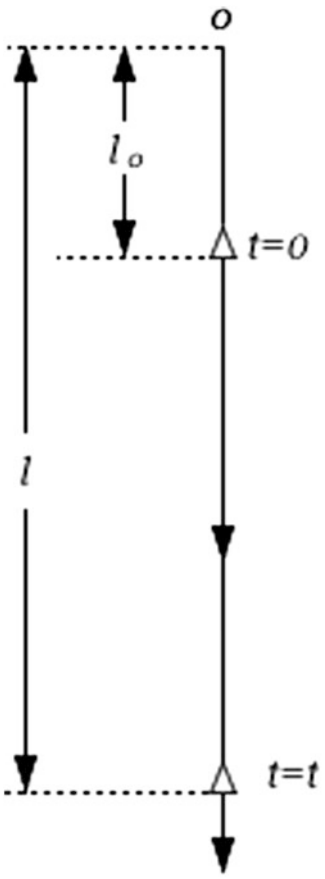


Fig. 2.18 Free-fall motion of an object

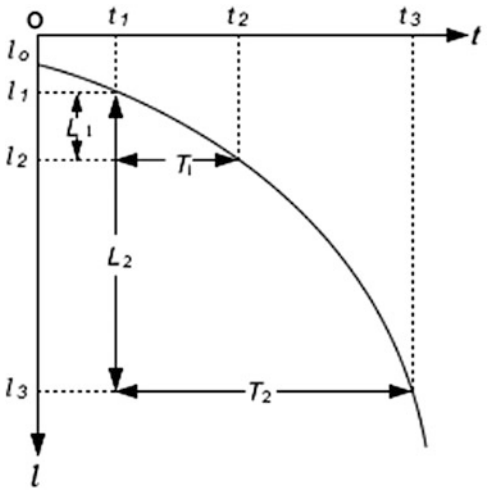


Fig. 2.19 Gravity determination by the free-fall method

$$l_1 = l_0 + V_0 t_1 + \frac{1}{2} g t_1^2,$$

$$l_2 = l_0 + V_0 t_2 + \frac{1}{2} g t_2^2, \text{and}$$

$$l_3 = l_0 + V_0 t_3 + \frac{1}{2} g t_3^2.$$

Subtracting the first equation from the second and third equations, respectively, the results are as follows:

$$\left.\begin{aligned} L_1 &= V_0 T_1 + \frac{1}{2} g T_1 (t_1 + t_2) \\ L_2 &= V_0 T_2 + \frac{1}{2} g T_2 (t_1 + t_3) \end{aligned}\right\}, \tag{2.26}$$

where $L_1 = l_2 - l_1$ and $L_2 = l_3 - l_1$ are the distances from the first position to the second and third positions, respectively. $T_1 = t_2 - t_1$ and $T_2 = t_3 - t_1$ are the times taken by the falling body in its motion from the first position to the second and third positions. To eliminate V_0, the two equations in (2.26) are divided by T_1 and T_2, respectively, and the two results thus achieved subtracted from each other, which reads as:

$$\frac{L_1}{T_1} - \frac{L_2}{T_2} = \frac{1}{2} g (t_2 - t_3).$$

Since $t_2 - t_3 = T_1 - T_2$, the formula of g can finally be written as:

$$g = \frac{2}{T_2 - T_1} \left(\frac{L_2}{T_2} - \frac{L_1}{T_1} \right). \tag{2.27}$$

Thus it can be seen that to determine gravity using the free-fall method requires knowledge of the distances L_1 and L_2 traveled by the falling body within the time periods T_1 and T_2.

Rise-and-Fall Method

In this method, an object is thrown vertically upward and then allowed to fall freely. To obtain the gravitational acceleration g, it is necessary to label two positions S_1 and S_2 in its course of motion. The time intervals T_1 and T_2 of the falling body past each position are determined, cf. Fig. 2.20. The transverse axis indicates time and the ordinate axis indicates the vertical position of the falling body. Let H_1 and H_2 be

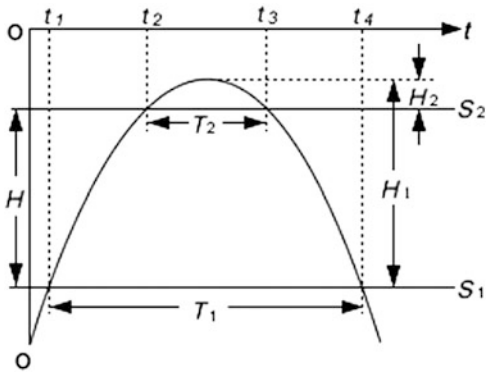


Fig. 2.20 Gravity determination by the rise-and-fall method

the distances from the two measuring positions to the peak of its motion. According to (2.25), where $l_0 = 0$, $V_0 = 0$, we will obtain:

$$H_1 = \frac{1}{2} g \left(\frac{T_1}{2} \right)^2,$$

and

$$H_2 = \frac{1}{2} g \left(\frac{T_2}{2} \right)^2.$$

H is taken to denote the distance between the two positions, and hence yields:

$$H = H_1 - H_2 = \frac{1}{2} g \left[\left(\frac{T_1}{2} \right)^2 - \left(\frac{T_2}{2} \right)^2 \right].$$

After rearrangement, the equation for g becomes:

$$g = \frac{8H}{T_1^2 - T_2^2}. \tag{2.28}$$

Thus, it can be seen that to determine gravity using the rise-and-fall method requires the determination of time intervals T_1 and T_2 of the object passing two positions with a distance of H during its rise and fall.

2.4.2 Relative Gravimetry

Relative gravimetry is a technique used to determine the gravity difference between two points, and then to obtain the gravity value of each point in a pointwise manner through at least one point of known gravity value.

The static method of relative gravimetry is to use a kind of force (such as the spring force) to work against the force of gravity that is acting on the object and balance the gravitational pull. By changing gravity, the location of the equilibrium position (location of the spring) is also changed. As long as the change of the equilibrium position (the amplitude of the spring) is determined, the variation in gravity can be calculated (according to Hooke's Law). The gravity difference between the two locations is thus obtained (Lu 1996).

Currently, the most frequently used gravimeter is called the spring gravimeter, in which the spring force is used to balance gravity. Examples are the quartz spring gravimeter ZSM series manufactured by Beijing Geological Instrument Factory and the (LCR) metal spring gravimeter by LaCoste and Romberg in the USA. Both of these spring gravimeters incorporate a spring mass system, optical system, mechanical device for measurement, instrument panel, and insulated case. The range of gravity difference measured by ZSM is 80×10^{-5} m/s^2 to 120×10^{-5} m/s^{2}and the precision of measurement is between 0.1×10^{-5} m/s^2 and 0.3×10^{-5} m/s^2 The LCR gravimeter can also be classified as Model G and Model D. The range of direct measurement of the Model G is up to $7{,}000 \times 10^{-5}$ m/s^2 It can be utilized for relative gravity measurement on a worldwide scale and its measurement precision amounts to $\pm 20 \times 10^{-8}$ m/s^2 The range of direct measurement of Model D is only 200×10^{-5} m/s^2 It is widely used in regional gravity surveys and its precision of measurement is slightly higher than that of Model G.

2.4.3 Airborne Gravimetry

Airborne gravimetry is a method employed to determine the near-Earth gravitational acceleration using an integrated airborne gravity remote sensing system, which consists of an aircraft as carrier, airborne gravimeter, GPS, altimeter, and attitude determination devices, etc. (Fig. 2.21). It can operate in areas where terrestrial gravity measurement is hard to conduct such as deserts, ice sheets, marshlands, and primeval forests. It can acquire information on the gravity field at a fast pace, with high precision, on a large scale, and with even distribution. Compared to the classical technique of terrestrial gravity measurement, it is entirely different in terms of measuring instrument, motion carrier, measuring technique, methods of data collection, as well as theory of data reduction, etc. Airborne gravity measurement has fully demonstrated the integrated application of modern technologies in the field of geodetic survey. It is of vital significance to geodesy, geophysics, oceanography, resources exploration, and space science.

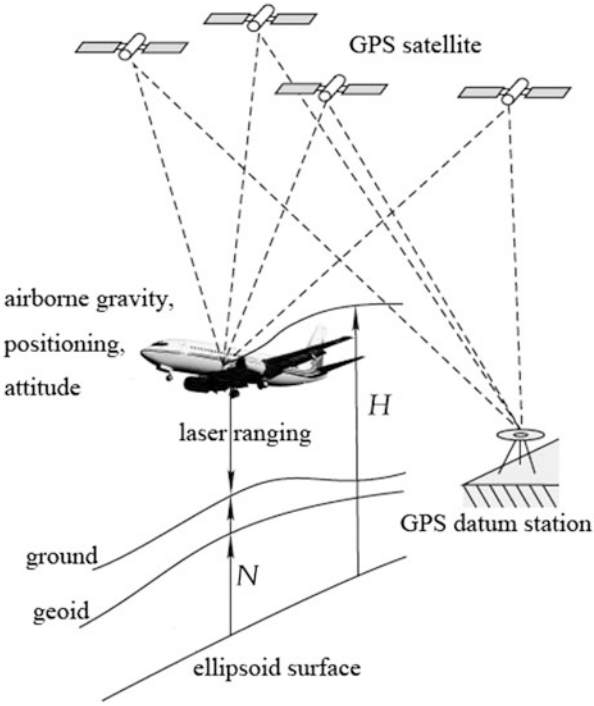


Fig. 2.21 Airborne gravity measurement

An airborne gravity measurement test was first conducted in 1958. The precision of navigation was rather low and the 10 mGal accuracy in the vertical disturbing accelerations of the aircraft was difficult to maintain, so until the late 1970s the technology of airborne gravimetry had virtually been in a state of stagnation. The advent of GPS, particularly implementation of the centimeter-level kinematic differential GPS, enabled the separation of gravitational effects with a precision of a few milligals. There are two main categories in airborne gravimetry; namely, scalar gravimetry and vector gravimetry. Scalar gravimetry can only determine the acceleration due to gravity, whereas vector gravimetry can measure both gravity anomalies and deflection of the vertical. Currently, the technology of airborne vector gravimetry is still undergoing research and development and is being used in some routine operations.

Fundamentals of Airborne Gravimetry

The basic principle of airborne gravimetry is to use the airborne gravimeter on the aircraft to determine the gravitational variations of the flight profile relative to the surface reference gravity point and compute the non-gravitational accelerations and corrections to disturbance. Through filtering and data processing, the results can be obtained and then, in the downward continuation approach, the gravity value at a

surface point can be obtained. Airborne gravimetry is relative gravity measurement, i.e., prior to taking off, the aircraft is connected to a surface point of known gravity. Its basic data model is:

$$\Delta g_h = g_b + \delta g - A_v - A_E - A_h + 0.3086H - \gamma_0, \tag{2.29}$$

where Δg_h is the gravity anomaly at a point in space at a height H, g_b is the gravity value at the ground gravity reference station, δg is the gravitational variation relative to g_b observed by the airborne gravimeter, A_v is the vertical acceleration correction of the aircraft, A_E is the Eötvös correction, A_h is the inclination correction to the horizontal acceleration, γ_0 denotes the normal gravity value (referred to in Sect. 4.1) evaluated on the geometric surface of the reference ellipsoid, and $0.3086H$ is the spatial correction of normal gravity.

The vertical disturbing acceleration for aircraft A_v is mainly induced by the vertical motion of the aircraft and the self-excited vibration in the body of the aircraft. This self-excited vibration is chiefly in the high-frequency bandwidth and can be removed by means of a low-pass filtering technique and high-damping of the gravimeter's sensing element. Vertical motion of the aircraft can be corrected by determining the flight altitude in progression with an appropriate computation method. It is fairly easy to measure changes in flight altitude relative to sea level, i.e., to measure directly the changes in distance from the aircraft to the sea surface using an altimeter. However, over land surfaces, what the altimeter measures are the changes in altitude from the aircraft to the ground; therefore, in order to obtain changes in the flight altitude, measurements of changes in the topographic surface of the predetermined flight course are also needed at the same time

To our knowledge, gravity is the resultant of the universal gravitation of the Earth's masses and the centrifugal force due to the Earth's rotation. When measuring gravity on a moving platform, the centrifugal force will change due to the resultant force of the carrier's velocity and the rotation velocity of the Earth, and this change is known as the Eötvös correction (A_{E}). The computational formula is written as:

$$A_{\mathrm{E}} = \left(1 + \frac{H}{R}\right)\left(2\omega V \sin A \cos\varphi + \frac{V^2}{R}\right), \tag{2.30}$$

where H denotes the flight altitude, R is the average radius of the Earth, V is the velocity of the carrier, A indicates the azimuth of the motion, ω is the angular velocity of the Earth's rotation, and φ is the geocentric latitude at a measuring point.

When determining gravity, the gravimeter and the level surface should be strictly parallel to each other. For airborne gravity measurement, if the platform of the gravimeter is not strictly parallel to the level surface, it will not only affect the gravitational acceleration but also exert influence on the vertical component of the horizontal acceleration. This effect is called inclination correction to the horizontal acceleration. Assuming that g is the actual gravity value, g_t is the value measured by the gravimeter, θ is the inclination between the platform surface and the level

surface, and A_e denotes horizontal acceleration, the inclination correction to the horizontal acceleration (A_{h}) can be expressed as:

$$A_h = g(\cos\theta - 1) + A_e \sin\theta. \tag{2.31}$$

In the above equation, when $A_e = 500$ mGal, $\theta \leq 3.\ 4'$, A_b is less than 1 mGal. Since the horizontal precision of the gyro platform is better than $0.\ 2'$, this correction is generally neglected and the corresponding error is less than 0.05 mGal.

System of Airborne Gravimetry

The airborne gravimetry system is the product of a combination of modern technologies like gravity sensing, satellite positioning, inertia and precision altimetry, etc. It mainly consists of five systems:

1. Gravity Sensor System. This mainly comprises the airborne gravimeter and the platform. The airborne gravimeter should have enough dynamic range and be able to provide information, like the large, short-time accelerations while the aircraft is taking off and landing, so as to facilitate the computation of corrections to the gravity disturbance.
2. Dynamic Positioning System. The major role of this system is to guarantee the optimal real-time navigation by using GPS, to provide data of initial orbit and precise location, and to compute accelerations in relation to the motion of the carriers. It is feasible to use merely pseudo-range measurement in real-time navigation. However, in order to obtain the precise flight path, it is necessary to integrate the utilization of pseudo-range measurement, phase measurement, and Doppler measurement to observe data.
3. Attitude Sensor System. The flight attitude of the aircraft is usually referred to as "pitch, roll, and yaw" and is determined by inertia measuring instruments. Because of the disadvantages that inertia measuring instruments are highly expensive, suffer from high drift, and are hard to maintain, in recent years GPS attitude measuring instruments with high precision, zero drift, and low price have come into use.
4. Altitude Sensor System. The major function of this system is to provide data on height for computing the Eötvös correction. The correction is applied to reduce the airborne gravity anomaly to the Earth's surface using microwave altimeter, radar altimeter, pressure altimeter, or GPS survey.
5. Data Collection and Processing System. This includes airborne data collection devices and ground data processing devices. The airborne devices are used to record the input data from the subsystems of the gravity sensor, navigation positioning, attitude sensor, and altitude sensor in synchronization. Each set of recorded data should have a unified accurate time scale to facilitate computation and processing for the ground devices.

2.4.4 Satellite Gravimetry

The features of satellite gravimetry chiefly include the ground tracking satellite, satellite-to-satellite tracking (SST), satellite gravity gradiometry (SGG), and satellite altimetry (SA) (see, e.g., Torge and Müller 2012). SA has been described in Sect. 2.3.

Determining the Earth's Gravity Field by Means of a Ground Tracking Satellite

The Earth's gravity field can be determined by means of a ground tracking satellite, using techniques such as satellite laser ranging (SLR), Doppler orbitography and radiopositioning integrated by satellite (DORIS), and precise range and range-rate equipment (PRARE).

The observed quantities involved with use of a ground tracking satellite include primarily the direction, range, range rate, and phase from the ground tracking station to the satellite being tracked. The geometric and physical functional relationships between the satellite orbit and the ground tracking station can be established on the basis of these observational data. Since the satellite orbit is the implicit function of the perturbation factors of the Earth's gravity field, the gravity field of the Earth can be computed.

Determining the Earth's Gravity Field by Means of Satellite-to-Satellite Tracking

The technologies of SST can be sorted into two modes: high–low satellite-to-satellite tracking (SST-hl) and low–low satellite-to-satellite tracking (SST-ll). SST-hl utilizes the space-borne GPS receiver and the GPS satellite constellation (altitude about 21,000 km) on the low Earth orbit satellite (LEO, altitude 400 km), forms the high–low satellite space tracking network, and estimates the three-dimensional location, velocity, and acceleration of the low Earth orbit satellite, namely the first derivative of gravitational potential (GPFD). SST-ll employs two identical satellites in the same orbit with an inter-satellite distance of 200–400 km, measures precisely the relative motion of the two satellites or the changes of inter-satellite distance by using a microwave interferometer, and determines the coefficients of the Earth's gravity field based on the rate of change of the inter-satellite distance (see Nin et al. 2006).

Germany's CHAMP (Challenging Mini-Satellite Payload for Geophysical Research and Application) employs the SST-hl tracking mode, as illustrated in Fig. 2.22. CHAMP was successfully launched into an orbit of 418–470 km altitude in July, 2000 on a 5-year mission. One of the scientific missions of CHAMP was to

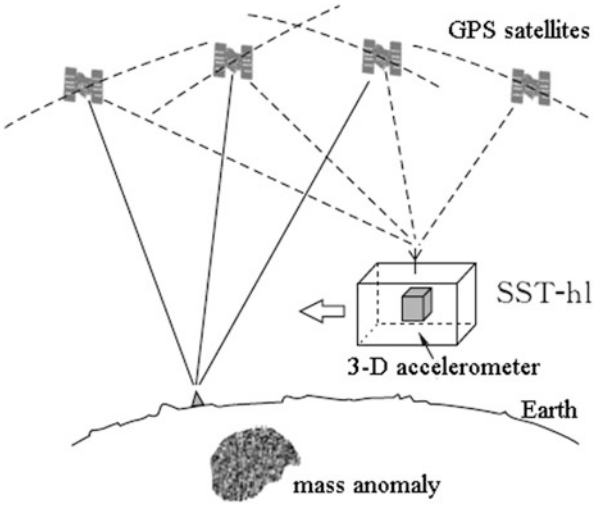


Fig. 2.22 Schematic diagram of SST-hl tracking mode

determine the medium- and long-wavelength static part and the temporal variations of the gravity field.

GRACE (Gravity Recovery and Climate Experiment) is a joint project of the USA and Germany. It employs the combination of two tracking modes: SST-hl and SST-ll, as shown in Fig. 2.23. The GRACE satellite was successfully launched into orbit in March, 2002 on a 5-year mission. One of the scientific missions of GRACE was to determine precisely the medium- and long-wavelength static part of the gravity field and to analyze and determine the variations in the Earth's gravity field every 2–4 weeks (see Klees et al. 2008).

Determining the Gravitational Acceleration Differences in the Earth's Gravity Field by Satellite Gravity Gradiometry

Satellite gravity gradiometry (SGG) allows determination of the differences in gravitational acceleration in three mutually orthogonal directions by using the differential accelerometer on one or more fixed baselines (about 70 cm) inside the satellite. The signals observed indicate gradients of the gravitational acceleration component, i.e., the second derivative of the gravitational potential. Non-gravitational accelerations (e.g., air resistance), in the same way, exert some effects on all the accelerometers inside the satellite, and the difference can be removed perfectly when differentiating. One of the missions of SGG is to detect the Earth's gravity field and its variations with higher temporal and spatial resolution.

The European Space Agency (ESA) launched a third gravity satellite in March, 2009, the GOCE (Gravity and Ocean Circular Exploration) with SGG mode (Fig. 2.24) (see Bouman et al. 2004).

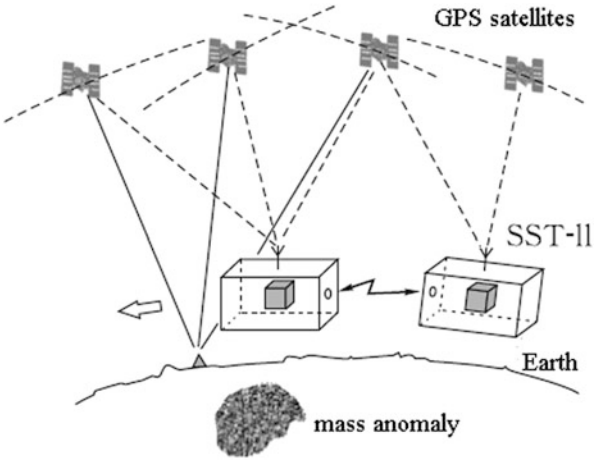


Fig. 2.23 Schematic diagram of SST-ll tracking mode

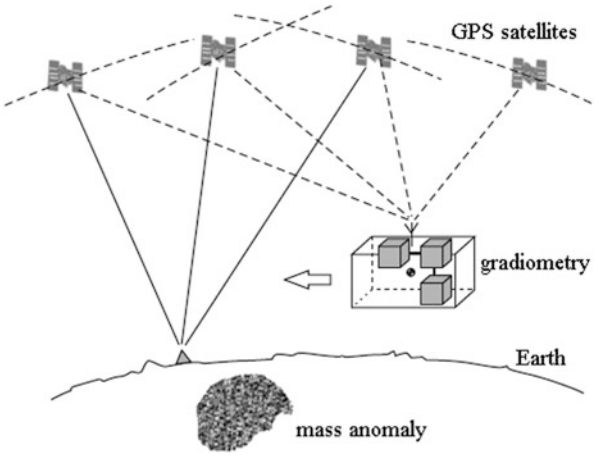


Fig. 2.24 Schematic diagram of SGG mode

Review and Study Questions

1. Briefly discuss the instruments employed and the observational quantities for different geodetic data collection techniques.
2. Briefly explain the concepts of horizontal and vertical angles and introduce the methods used for measuring horizontal angles.
3. Illustrate the concepts of astronomical longitude, astronomical latitude, and astronomical azimuth.
4. Briefly describe the basic principles of and methods for electromagnetic distance measurement (EDM).
5. Briefly discuss the methods for GPS measurements.

6. What are the basic principles for determining marine geoid height using satellite altimetry?
7. Explain the measuring principles of the VLBI technique.
8. How many categories can satellite gravimetry be classified into and what are they?
9. Explain the advantages of space geodetic techniques with respect to the classical geodetic survey.

Chapter 3
Geodetic Datum and Geodetic Control Networks

To measure terrain, surface features, position coordinates, heights, and gravity values at points on the Earth's surface, there need to be corresponding reference points or surfaces (also known as datum points or surfaces), namely geodetic datums, to which surveying and mapping results are referred. Geodetic datums consist chiefly of coordinate datums (including classical horizontal datums and three-dimensional coordinate datums), vertical datums, sounding datums, as well as gravity datums. Geodetic datums provide initial data for all kinds of surveying and mapping work and serve as the foundation for determining the geometric shape and spatial–temporal distribution of geospatial information. Again, it is geodetic datums that are referred to when the spatial positions of geographical features in the real world are expressed in the data space. The missions of constructing geodetic datums include determining and defining the coordinate system, height system, and gravity reference system, and establishing and maintaining the coordinate framework (horizontal and satellite geodetic control networks), elevation framework (vertical control network), and gravimetric framework (gravity control network).

Classical horizontal and vertical datums are realized by classical geodetic methods. Due to their limited controlling area, these two datums can only be used as regional datums and are usually applicable countrywide. The three-dimensional coordinate datums and gravity datums can be used as both global and regional datums. The datums are represented by the position coordinates, heights, and gravity values at a series of control points. To be specific, the datums are realized through establishing different geodetic control networks, i.e., the extensions of horizontal and vertical datums, three-dimensional coordinate datums, and gravity datums are realized by the horizontal and vertical control networks, satellite geodetic control networks, and gravity control networks, respectively.

The horizontal coordinates and heights of points on the Earth's surface determined by classical geodetic techniques refer to different datums. For instance, methods such as triangulation and traversing can only obtain the two-dimensional horizontal coordinates (x, y) or (L, B) of the surface point, whereas height H of the point can only be obtained by leveling or trigonometric leveling. Because the

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_3,

principles and methods for determining the horizontal coordinates and heights are fundamentally different, we cannot use the same datums in determination. Conventionally, the establishment of geodetic control networks, as a consequence, separates the horizontal from the vertical control network. The horizontal control network determines the horizontal datum for a surface point whereas the vertical control network determines the vertical datum for a surface point. Such control networks established through two suites of systems are also called a "2 + 1" network, which is still widely used in operations. In modern geodetic survey, the horizontal control network is often established by GPS methods.

This chapter mainly discusses geodetic datums and the methods, principles, and plans for establishing geodetic control networks.

3.1 The Horizontal Datum and Horizontal Control Networks

The horizontal datum is realized by horizontal coordinates of a series of geodetic control points, usually through establishing horizontal control networks. It serves as the basis for determining the horizontal coordinates of points on the Earth's surface.

3.1.1 Geodetic Origin and the Horizontal Datum

Geodetic Origin

The geodetic origin is the initial or starting point for computing geodetic coordinates in the national horizontal control network. It is chosen as an appropriate but arbitrary point in the national geodetic network. Geodetic reference data are obtained by determining with high precision the astronomic longitude and latitude of the geodetic origin and the astronomical azimuth from there to another point; and finding the geodetic longitude, latitude, and height of the geodetic origin and the geodetic azimuth from there to another point according to the principle of ellipsoid orientation. The reference ellipsoid and ellipsoid orientation will be discussed in Chap. 7.

The geodetic origin of China Beijing Geodetic Coordinate System 1954 is located at the Pulkovo Astronomical Observatory of the former Soviet Union. In the 1970s, China decided to build its own independent geodetic coordinate system. After field survey and comprehensive analysis, the geodetic origin of P.R. China was at last established in Shijisi Village, Yongle Township, Jingyang County, Shaanxi Province, with the coordinate longitude of E 108°55′25.00″and latitude N 34°32′27.00″, and it was completed in 1978 (see Chen 2003). China's geodetic origin consists of the survey mark at the center, the instrument platform, the main

Fig. 3.1 Outside appearance of China's geodetic origin. *Source*: SBSM

building, and the projection platform. The instrument platform was set in the observation room on the top floor of the seven-storey building which is 26 m above the ground. The top of the building is a semicircular dome made of fiber-reinforced plastic (FRP) and automatically controlled by electric power for celestial observation (cf. Fig. 3.1). The survey mark at the center of the geodetic origin is set in the basement center of the main building, which is made of red agate with a diameter of 10 cm, and is delicate and durable (Fig. 3.2).

The Horizontal Datum

The horizontal datum provides the basis for establishing the national geodetic coordinate system and for calculating the geodetic coordinates of each point in the national horizontal control network. It includes a set of initial data, i.e., the geodetic longitude and latitude of the initial point and the geodetic azimuth from the initial point to its adjacent point in the national geodetic control network. (The initial point is the geodetic origin in classical geodetic survey.)

The extension of the horizontal datum is realized by the horizontal control network formed by a series of control points. Coordinates of the control points are computed from the geodetic origin and obtained by classical geodetic methods such as traversing, triangulation, and so on. In modern geodetic survey, the horizontal datum is usually realized by 3-D datum obtained from the GPS method (cf. Sect. 3.3).

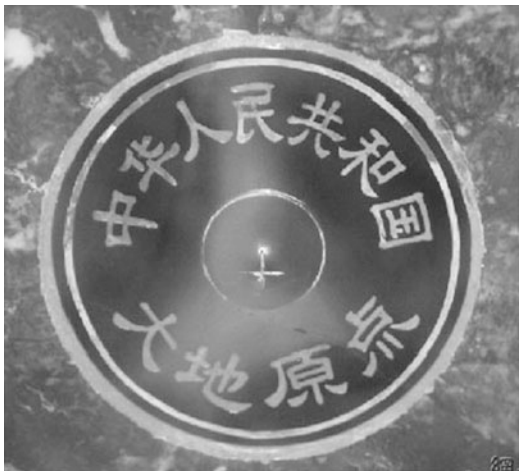


Fig. 3.2 Geodetic origin mark of China's classical geodetic control network. *Source*: SBSM

3.1.2 Methods of Establishing a Horizontal Control Network

Traversing

A series of geodetic control points P_1, P_2, P_3. . ., each intervisible with its adjacent points, are chosen to form a system of broken lines called a traverse, as shown in Fig. 3.3. The distances between nearby traverse points and the angles at all traverse points are measured and reduced to a plane. Assume $D_{12}, D_{23}, \ldots$ are the lengths of the traverse lines in the plane, and β_i is the horizontal angle at each traverse point; knowing the plane coordinates (x_1, y_1) of point P_1 and the grid bearing (grid azimuth) T_{10} of P_1P_0, the grid bearing of each traverse leg can be obtained starting from T_{10}, namely:

$$\begin{aligned} T_{12} &= T_{10} + \beta_1, \\ T_{23} &= T_{12} + 180^\circ + \beta_2. \\ &\ldots \end{aligned}$$

According to these bearings and lengths of the traverse lines, the coordinates of other traverse points can be obtained from the coordinates of point P_1:

$$\begin{aligned} P_2: \quad & x_2 = x_1 + D_{12} \cdot \cos T_{12}, \\ & y_2 = y_1 + D_{12} \cdot \sin T_{12}. \\ P_3: \quad & x_3 = x_2 + D_{23} \cdot \cos T_{23}, \\ & y_3 = y_2 + D_{23} \cdot \sin T_{23}. \\ & \ldots \end{aligned}$$

This is the fundamental principle of establishing the horizontal control network by traversing.

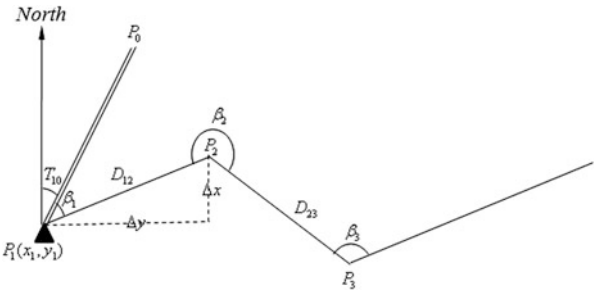


Fig. 3.3 Traversing

Triangulation

A series of points P_1, P_2, P_3. . ., each point intervisible with its adjacent points, are chosen and connected in the form of a triangulation network, as shown in Fig. 3.4. Determine the length of P_1P_2 and its azimuth as the initial length and azimuth of the network, respectively; observe the angles in each triangle and reduce these lengths and angles to a plane. Assume the given coordinates of point P_1 are (X_1, Y_1), the length and grid bearing of P_1P_2 are D_{12} and T_{12}, respectively, and the observed angles are A_i, B_i, and C_i. The lengths and grid bearings of each side can be obtained from P_1P_2:

$$D_{13} = D_{12}\frac{\sin B_1}{\sin A_1}, \quad T_{13} = T_{12} + C_1.$$

$$D_{14} = D_{13}\frac{\sin B_2}{\sin A_2}, \quad T_{14} = T_{13} + C_2.$$

$$\ldots$$

The coordinates of points in the entire network can be obtained according to these side lengths and azimuths, namely:

$$x_3 = x_1 + \Delta x_{13} = x_1 + D_{13}\cos T_{13},$$
$$y_3 = y_1 + \Delta y_{13} = y_1 + D_{13}\sin T_{13}.$$
$$\ldots$$

This is the fundamental principle of establishing the horizontal control network by triangulation.

Trilateration and Combination of Triangulation and Trilateration

Establishing the horizontal control network using trilateration is almost the same as when using triangulation. The difference is that while using trilateration we only measure the side lengths of the triangle in the network. The interior angles are obtained through computation. If some or all lengths are also measured apart from

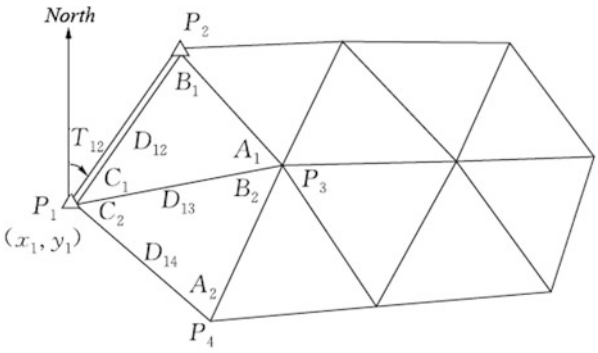


Fig. 3.4 Triangulation

measuring the angles, it will be called the method of combined triangulation and trilateration or triangulateration.

3.1.3 Principles of Establishing a National Horizontal Control Network

A national horizontal control network is a fundamental construction project. We therefore need to proceed from the real situations of a nation, properly handle the relationship between quality, quantity, time, and expenditure according to theory and the real experience of network establishment, and work out specific principles as the basis for designing and establishing the geodetic control networks.

Network Establishment and Control Based on Hierarchical Orders

A national horizontal control network can be established based either on a single order or several different orders. The single-order control network can serve directly as the basis for mapping control and is usually established in countries with smaller territories to ensure more homogeneous accuracy throughout the network and facilitate adjustment computations. Countries with vast territories often adopt the method of establishing networks from higher to lower orders. They usually first build a nationwide primary control network with higher accuracy and lower density as a consistent control framework, and then continue to densify the control network in a piecemeal fashion according to the needs of different areas. The side lengths of control networks become shorter and the accuracies get lower as the order changes from higher to lower. Using such a method to establish other triangulation networks successively in different areas within one consistent coordinate system can not only satisfy the desired accuracy but also achieve the effective results at a faster pace and lower cost.

The national horizontal control network in China is classified into four orders. The first-order triangulation chain with high accuracy and low density crisscrosses the whole country to form the key network of the unified coordinate system. Then, the second-, third-, and fourth-order horizontal control networks are established in a piecemeal manner according to actual needs. Special requirements are with respect to the following factors.

Sufficient Accuracy

Apart from being the control framework of the national unified coordinates, the first- and second-order networks, in the process of establishing the national horizontal control network, have to meet the requirements for mapping of the basic scale topographic maps and the development of modern technology, such as space technology, precise engineering, earthquake monitoring, and geodynamics, whereas the third- and fourth-order horizontal control networks are used chiefly for a higher-level control of the topographic mapping control points and to satisfy the needs of fundamental engineering construction. Control points of various orders, therefore, must cater for the actual demands. For example, the accuracy of the first- and second-order control points should meet the needs of a 1:50,000 scale topographic map, while that of the third- and fourth-order control points should meet the needs of topographic mapping at a scale of 1:10,000.

Necessary Density

Density of the control points in the control network means that there is usually one single point every several square kilometers on average. It can also be expressed by the average side length of midpoints in the control network. The shorter the side length, the denser the geodetic points will be. The controlling area Q of each point is expressed by the average side length S, namely:

$$Q = S^2,$$

and

$$S = \sqrt{Q}, \tag{3.1}$$

which is the relationship between the side length and the controlling area.

The density of the points is required to be different according to different mapping scales and methods. On average, three or four geodetic points are generally required to densify control points for each map sheet. For different engineering projects, however, the desired density of points will presumably be different and should be determined according to real situations.

Unified Specifications and Standards

China has a relatively large territory, so it costs a great deal of manpower, material, and financial resources to build the national control network, which has to be accomplished with the concerted efforts of many departments and agencies. Once established, some parts of the geodetic network still need to be reinforced, upgraded, and improved according to the needs of different sectors. To avoid redundancy and waste of resources and to encourage sharing of benefits, the work plan for network establishment and the specific operational standards should be unified so that the accuracies of the results obtained by different surveying and mapping institutions and the specifications of design meet the requirements and such results will become an integral unity as a component of the national control network. Issues like the overall plan, the anticipated target of precision, datum selection, and so on should be in compliance with the technical standards. In addition, such issues as specific implementation plans, instrument utilization, operation methods, tolerance restrictions, and project checking and acceptance should also be provided in the specifications.

3.1.4 Plans for Establishing a National Control Network

The China national horizontal control network built in the 1950s used primarily the method of triangulation, but traversing was also employed in harsh areas. Below is an overview of the establishment of the triangulation chains of various orders.

First-Order Triangulation Chain

The first-order triangulation chain is a national primary network, used to build a precise framework of a unified coordinate system throughout the country to control the establishment of the second- and lower-order triangulation networks and provide data for studying the size and shape of the Earth and geodynamics. Mapping control is not the direct objective—accuracy has more importance in this case.

The first-order triangulation chain runs along the meridian and the parallel as shown in Fig. 3.5. The triangulation chain between the intersections is called the chain section; the circle formed by the east–west and north–south chain sections is called the chain loop; many chain loops form the chain system. The chain section is approximately 200 km long and is usually formed by single triangles and may also include some geodetic quadrilaterals or mid-point polygons. The average side length of triangles in the chain ranges from 20 to 25 km, any arbitrary angle of triangles is not less than 40°, and the distance angles of the geodetic quadrilaterals or mid-point polygons should be greater than 30°. Computed by the triangle

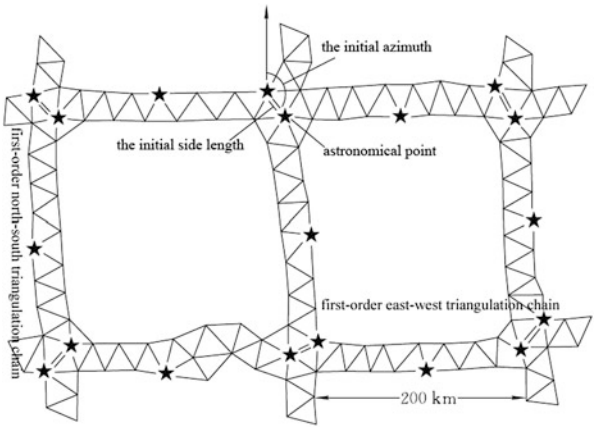


Fig. 3.5 First-order triangulation chain with astronomical points

closure, the mean square error of angle observation should not be greater than $\pm$ 0. 7″.

The initial side length at the crossing of the chain sections should be determined with a relative accuracy of no less than 1/350,000. The astronomical longitude, latitude, and azimuth are measured at the two endpoints of the initial side and the former two are also measured at a point in the center of the chain. The determined mean square error of the astronomical longitude, latitude, and azimuth should be less than $\pm 0.3''$, $\pm 0.3''$, and $\pm 0.5''$, respectively. All points with measured astronomical longitude and latitude will provide data for computation of the deflection of vertical. As astronomical surveying is involved in the plans for network establishment, the national horizontal control network is also called the astro-geodetic network.

Second-Order Triangulation Network

Set within the area circled by the first-order triangulation chain loop, the second-order network is the overall basis for densification of the third- and fourth-order networks, as shown in Fig. 3.6. The average side length of the second-order network is 13 km and the density of such a network basically satisfies the needs of the 1:50,000 scale mapping. The second-order network, together with the first-order chain, belongs to the national high-order network. Hence, accuracy should be the primary concern whereas density is secondary. The mean square error of angle observation computed through the triangle closure should be less than $\pm 1''$. An initial side and azimuth are to be determined at the center of the network. For larger chain loops, the initial azimuth should be measured as well. Angles of triangles in the network should be no less than 30°. The second-order network on either side of the first-order triangulation chain should be connected with the first-order chain to form a continuous triangulation network.

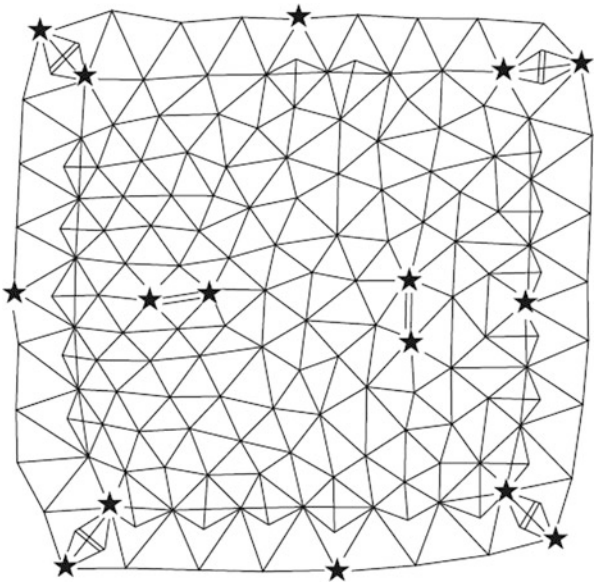


Fig. 3.6 Second-order continuous network

Third- and Fourth-Order Triangulation Networks (Points)

National third- and fourth-order triangulation networks (points) can be further densified on the basis of the second-order network, as illustrated in Figs. 3.7 and 3.8. They are foundational to the mapping control survey and their density should accord with the mapping scale. The average side length of the third-order triangulation network is 8 km and the controlling area of each point is roughly 50 km^2, which can basically meet the needs of 1:25,000 scale mapping. The average side length of the fourth-order network is 4 km and the controlling area of each point is around 20 km^2, which can meet the needs of 1:10,000 and 1:50,000 scale mapping.

At each point of the third- and fourth-order networks there will be stations set for observation. The mean square error of angle observation computed through triangle closure should be less than $\pm 1.8''$ and $\pm 2.5''$ for the third- and fourth-order networks, respectively.

Traverse Control Network

Although traversing is not as effective as triangulation in controlling area, checking conditions, and constraining error propagation of azimuths, it still has distinct advantages that allow the network to be established in a flexible manner. Meanwhile, the survey can easily be carried out and terrain obstacles are more likely to be overcome. In the early 1960s, the first- and second-order control networks were sparsely established by traversing in most areas of the Tibetan Plateau. Traversing has become increasingly widely used with the prevalence of total stations and the

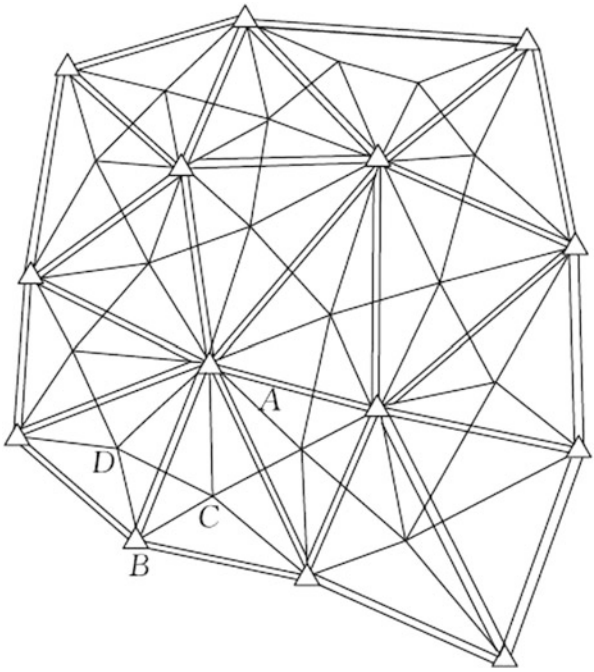


Fig. 3.7 Network densifying through point inserting

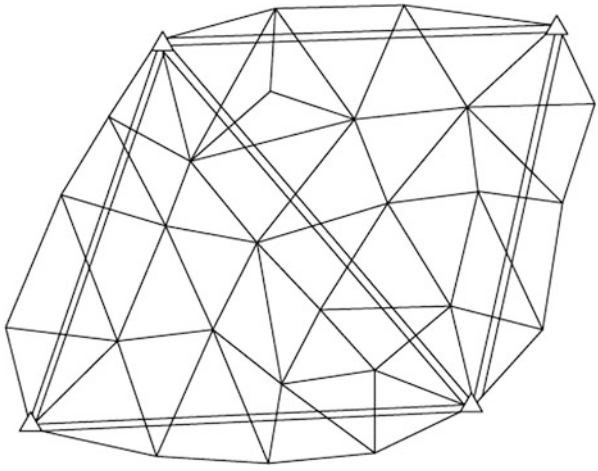


Fig. 3.8 Network densifying through sub-network inserting

constant advances in electromagnetic distance measuring (EDM) instruments to improve their accuracy, increase the measurement range, and lighten their weight. Traversing priorities are in densifying lower-order control networks and in replacing the third- and fourth-order triangulation networks to control large-scale mapping. Traversing is again a preferred method for connection survey in the military battlefield.

The establishment principle of traversing is fundamentally the same as that of triangulation and can also be classified into four orders. The accuracies required for the measured distances and angles of various-order traverses should generally accord with the accuracies obtained from computations by the triangulation chain of the corresponding order.

The first-order traverse usually runs along the main traffic routes and criss-crosses into larger traverse loops; several traverse loops constitute the traverse network. Figure 3.9 shows a schematic diagram of the traverse layout. The first-order traverse network should be properly connected with its adjacent triangulation chain to form an integral geodetic control network. The circumference of the first-order traverse loop is usually around 1,000–2,000 km. The second-order traverse is set within the first-order traverse loop (triangulation chain) with its two ends closed at the first-order traverse points (or triangulation chain), in the form of a connecting line. Simultaneously, mutually intersecting traverse loops are also formed between the second-order traverses, which constitute a network. The circumference of the second-order traverse loop is usually 500–1,000 km or so. The lengths of the first- and second-order traverse legs vary from 10 km to 30 km. To control the azimuth errors of the traverse legs and reduce the lateral errors of traverses, the astronomical longitude, latitude, and azimuth at the two ending points of each traverse leg should be measured in order to determine the initial azimuth of this traverse leg at a distance interval of 100–150 km from the first- and second-order traverses, where the first- and second-order traverses meet the first- and second-order triangulation chains and where the first- and second-order traverses intersect, so that the initial azimuth of this line in question can be obtained. Traverse is not as durable as a triangulation networks and the error propagation in calculating azimuths accumulates fast; the intervals of the initial azimuth should therefore be small. The traverse controlled by azimuths at the two ends is called the traverse section, which should be laid straight. Traverse section that generally runs along the same direction in the traverse network between two intersections is called a traverse segment. The propagation of errors in azimuth increases with the increasing of number of sides. Therefore, the legs of each traverse section in the first- and second-order traverses should be no more than seven.

The third- and fourth-order traverses are densified based on the first- and second-order traverse networks (triangulation chains), which are closed on a second known point, called the connecting traverse. The total length of such a single connecting traverse should be less than 200 km and 150 km for the third- and fourth-order networks, respectively. When several such connecting traverses are established, it would be preferable that these traverses constitute a network to reinforce the traverse's structure. The lengths of the third- and fourth-order traverse legs can be determined according to the performance of the side and angle measuring instruments and the density required for geodetic points. Generally, the lengths of the third-order traverse legs range from 7 to 20 km, while those of the fourth-order range from 4 to 15 km. A traverse leg with greater length is preferred in fieldwork.

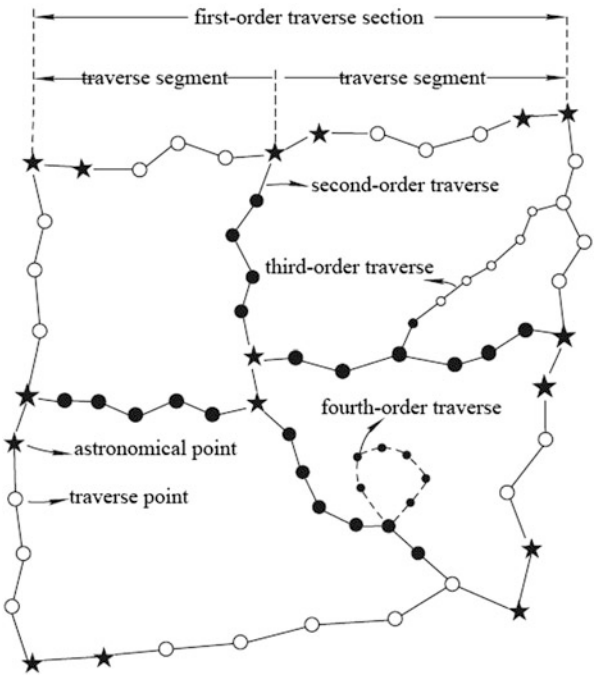


Fig. 3.9 Traverse network

Overview of China's Astro-Geodetic Network

The first-order triangulation chain together with the second-order network in China is called the astro-geodetic network, which started to be built in 1951 and was completed in 1971. The entire length of the first-order triangulation chain is roughly 80,000 km, including over 400 chain segments that have formed more than 100 chain loops and a total of over 5,000 first-order triangulation points (trig points); cf. Fig. 3.10. The overall adjustment of the astro-geodetic network was completed in 1982. Such an astro-geodetic network consists of the first-order triangulation chain, second-order network, some third-order network, as well as traverses. There are nearly 50,000 control points, 467 initial sides, and 916 initial azimuths. Approximately 300,000 error equations and more than 150,000 normal equations are formed. The adjustment shows that the mean square error of the point farthest away from the geodetic origin is ± 0.8 m and mostly the relative accuracy of adjacent points is less than 1/200,000.

3.1.5 Establishment of a Horizontal Control Network

Establishment of the horizontal control network includes technical design, reconnaissance for site selection, erection of survey marks, monument setting, distance

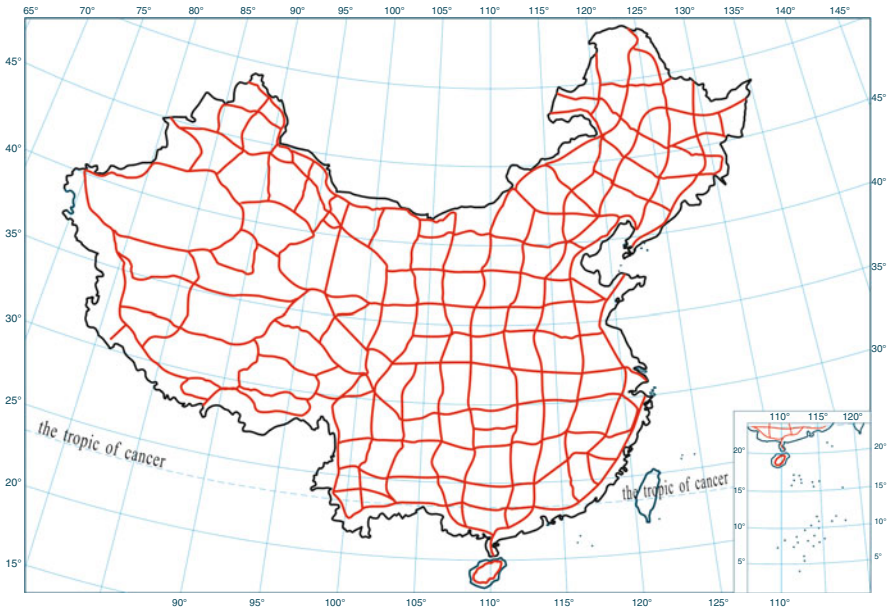


Fig. 3.10 China astro-geodetic network: *curves* indicate the first-order triangulation chain; *blank areas* in the *curved quadrangles* indicate the second-order triangulation network

measurement, angular measurement, adjustment computations, etc. (see SBSM 2004). Our discussions will focus on the first four aspects.

Requirements for the Position of Control Points

The position of horizontal control points should satisfy the following requirements for either technical design or reconnaissance for control point selection:

1. The side lengths, angles, and graphical structures formed between control points should completely conform to the requirements in the corresponding technical standards.
2. The control points should be marked where the sites can be extended easily and lower-order points are conveniently densified.
3. The position should be selected where the survey mark can be well preserved over time and it will be safe and convenient to erect the monument and to observe it. Therefore, the position should be selected in high land with solid soil and a fine drainage system, and should be a suitable distance away from highways, railways, high-voltage wires, and other buildings.
4. The line of sight should avoid slopes or coastlines of lakes and rivers to ensure a clear and stable image of the observation target and to reduce the effects of horizontal refraction in order to improve the accuracy of observation. The line of

sight should go beyond or deviate from obstacles by a certain distance which, for first- and second-order, respectively, should be no less than 4 m and 2 m in mountainous areas while no less than 6 m and 4 m in plain areas.

Technical Design

Data Collection

Data relevant to the survey areas should be collected before planning, including maps of various scales, aerial photo maps, traffic maps and meteorological information, existing results of geodetic points, natural and social geographical environments of the survey areas, transportation and material supplies, and so on. These data should be analyzed and studied as the basis and reference for the technical design.

Drawing Up Designs

Drawing up designs is a key aspect in technical design that deserves careful consideration in order to facilitate site selection. Fieldwork will otherwise be difficult.

Drawing up designs usually follows the steps and methods listed below:

1. Splicing the 1/50,000 or 1/100,000 scale topographic maps of the survey area and marking the already established triangulation chains, GPS networks, traverse networks, and leveling lines on the map.
2. Extending outward from the points of known control in a pointwise manner according to the requirements for positions of control points while considering creating the best figure possible. The points are laid out from higher to lower orders, from points of known control to unknown control, and from the interior to exterior in a pointwise fashion.
3. Drawing up the leveling connection lines according to the density requirements for the zero elevation surface provided in the corresponding technical standards; Trying to utilize the old network points already existing and to propose the plans for a connection survey.
4. Ensuring intervisibility during the site selection. Several alternative plans should be drawn up for the uncertain positions of points or directions.

Reconnaissance for Site Selection

The task of reconnaissance for site selection is to put the designs on the map into practice. The most appropriate site available should be selected according to requirements for the position of control points.

Erection of Survey Marks and Monument Setting

Erection of Survey Marks

National triangulation points or traverse points are way apart from each other and invisible in general conditions; geodetic survey marks are therefore needed to show the specific position of the point as the target. Geodetic survey marks are usually classified into ordinary survey marks and tower structures. The former is only used as the target point with a height of 4.3 m and 6.3 m, while towers are chiefly used to elevate the instrument and position of the target point when two adjacent points are not intervisible.

Monument Setting

The survey mark is a permanent mark of the control point position. Field observation is referred to the center of the mark, and the plane coordinates and height of the point obtained will be the position of the survey mark center. If the mark is destroyed or displaced, the surveying results and position coordinates will be meaningless. Hence, when the monument is set, one should firmly abide by the principle of "quality first." Meanwhile, the monument should be stable to ensure permanence.

A monument can be classified as that of the first- and second-order triangulation (traverse) points or that of the third- and fourth-order triangulation (traverse) points. A monument is generally filled with concrete chiseled from granite, bluestone, or other hard stones with identical specifications. Monuments consist of disks and pillars, both with a mark sunken into the center of their top surfaces. The survey mark can be made of metal or vitreous enamel. There are many types of monuments, which are different in terms of the different orders and places of monumentation under the principle of ensuring their stability and permanence. Generally, a monument of first- and second-order points is composed of pillars and upper and lower disks, as shown in Fig. 3.11, while that of third- and fourth-order points is composed of pillars and one disk.

Completion of the technical design and erection of survey marks and monuments marks the position of each control point in the horizontal control network on the Earth's surface. However, extensive distance and angle measurements, as well as adjustment computations, still need to be made before the coordinates of the control points can be determined.

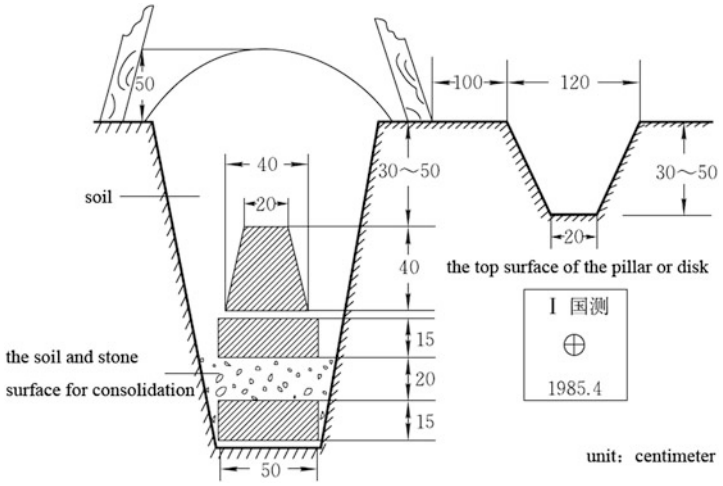


Fig. 3.11 Monumentation of the first- and second-order trig points

3.2 The Vertical Datum and Vertical Control Networks

A vertical datum (height datum) is realized by a vertical control network that provides the basis for determining the height of a point on the Earth's surface. The vertical control network is also known as the leveling network, which is established using leveling as the primary method, supplemented by trigonometric leveling. Leveling can be classified into four orders: the first- and second-order leveling is known as precise leveling, and the third- and fourth-order leveling is referred to as ordinary leveling. Correspondingly, the vertical control network can also be divided into four orders.

3.2.1 The Vertical Datum and Leveling Origin

A vertical datum provides the reference surface relative to which heights are measured. All leveling heights in the national unified vertical control network are calculated and determined with respect to the vertical datum, which consists of a zero elevation surface and a permanent leveling origin. Theoretically, the geoid is usually used as the reference surface for heights, yet practically the mean sea level (MSL) determined by averaging the level of water at a tide gauge over time is often used as the level surface to which heights are referred. Overwhelmingly, the majorities of countries and areas worldwide have chosen the MSL to be the reference surface as this average position actually exists, is very stable, and can be determined precisely. Another benefit is that the global MSL also approximates to the physical surface of the Earth.

To determine the MSL, a tidal station will usually be established at the coast where it will be appropriate to record sea levels over time. The sea level changes due to changing external conditions and factors from inside the Earth, particularly the variations in positions of the Moon and the Sun. Such variations are periodic and the long period is the astronomical tide cycle, which spans approximately 18.61 years. Statistics have shown that the long periodic average sea level is by and large invariable and can be considered as the MSL for a particular area. From analysis of long-term sea level data we can see that there has been a several centimeter increase in global mean sea level each year, resulting from the rising temperature over the last 100 years. Two main factors have contributed to observed sea level rises. The first is thermal expansion of ocean water. The second is the contribution of land-based ice due to increased melting of glaciers and ice sheets. Therefore, corrections need to be applied to the vertical datum because of the long-term changes in mean sea level.

As the level surface to which heights are referred, MSL is often defined as zero height for a local area, also known as the zero point. To mark the position of the reference surface for heights (zero point) with clarity and stability, one also needs to establish a permanent benchmark and connect it to the MSL by precise leveling as the reference point for height measurements in a national or local control network. Such a benchmark is known as the leveling origin.

China's leveling origin is located at Qingdao Observatory Hill. One origin and five subordinate points constitute the leveling origin network. The subordinate points are used to monitor stability of the origin and to secure precision of the leveling connection. The leveling origin is established in a stone house to ensure its security. The inside of the house is built of granite from Mount Lao. At the center and four corners of the rooftop are five stone columns. The stone columns are exquisite and unique, chiseled and carved with delicate touches. The stone house covers an area of 7.8 m^2. Outside the stone house are very high double stockade fences and, inside, the doors are locked fast with threefold locks. The architecture is of Russian style and was built in 1954 (cf. Fig. 3.12). The interior wall is inlaid with a black marble stele on which the inscription "Leveling Origin of People's Republic of China" is engraved. Within the house there is a dry well roughly 2 m deep. At the bottom of the well is a priceless yellow agate as big as a fist. On the agate identifier is a double-layered cap made of brass and stone. A little red dot on the agate is the mark of the leveling origin (Fig. 3.13).

In China, the Huanghai (Yellow Sea) MSL obtained from the tidal station at the west end of pier No. 1 in Qingdao Large Port is taken as the reference surface from which heights are measured. The tidal station is geographically located at 120°18′40″ east longitude and 36°05′15″ north latitude. Inside, there is a tide-gauge well with a diameter of 1 m at a depth of 10 m. Three inlet tubes 60 cm in diameter have direct contact with the open sea. The tide gauge originally used was the German-made float automatic tide gauge and the tidal records began in 1900. It was destroyed during World War II. In 1947, the tide gauge was renewed and resumed work. After the founding of P.R. China the building was renovated and the facilities upgraded. The instruments currently in use are the HCJ1 tide gauges (also

Fig. 3.12 Outside appearance of the leveling origin in China. *Source*: SBSM

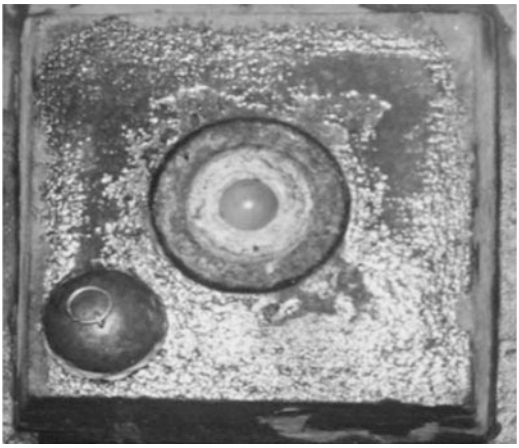


Fig. 3.13 Surface of the leveling origin mark in China. *Source*: SBSM

known as Valdai), the US imported SUTRON9000 automatic water gauges, and the SCA6-1 acoustic water level meter produced by the Research Institute of Technology, State Oceanic Administration. The water levels are observed three times a day at 07:45–08:00, 13:45–14:00, and 19:45–20:00. The tidal records obtained over years are measured and analyzed rigorously to calculate the sea level at the Qingdao tide gauge station. It is 2.429 m, which is considered the national vertical datum.

In 1959, China promulgated the *Guidelines for Geodetic Surveying in People's Republic of China* prescribing that "the height of the national benchmark is calculated relative to the leveling origin at Qingdao. According to the result computed in 1956, the height of the origin was designated 72.289 m above the Huanghai mean sea level." This reference surface is generally called the "Huanghai Mean Sea Level 1956."

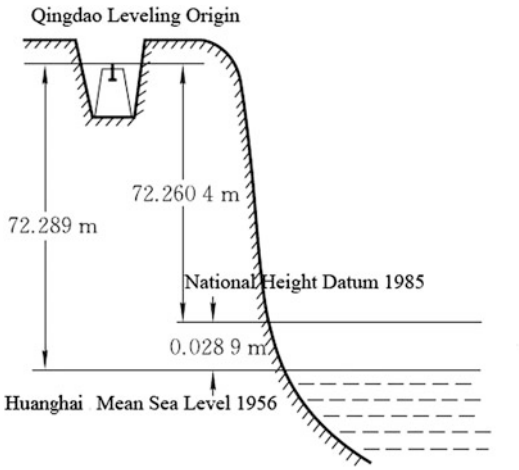


Fig. 3.14 Relations between different height datums and the leveling origin in China

The data used to calculate the "Huanghai Mean Sea Level 1956" was obtained from tidal observations between 1950 and 1956. As the time span was rather short, the result was far from satisfactory. In May 1987, China implemented the "National Height Datum 1985." This new datum used the tidal observations from 1952 to 1979 at Qingdao tidal station, and the Huanghai MSL was obtained by averaging heights of the sea levels in this area. The height of the leveling origin is 72.2604 m by precise leveling connection, differing from the Huanghai Mean Sea Level 1956 by 0.0289 m (cf. Fig. 3.14).

3.2.2 The Sounding Datum

Concept of the Sounding Datum

In the domain of marine surveying and mapping, the concepts of depths and the sounding datum are needed for describing seafloor surface features and carrying out corresponding bathymetric surveys. Depth is the sounding value obtained through bathymetric surveys of oceans and coastal sea waters starting from the sea surface at the time of measurement (i.e., instantaneous sea surface). Due to such effects as tides, sea waves, sea currents, and so on, the instantaneous sea surface height is subject to temporal variations. Hence, the instantaneous water depth obtained at the same sounding point at a different time also varies. Therefore, a fixed water surface should be specified as the reference surface for depth measurement. All the soundings measured at different times should be reduced to this reference surface, which is known as the sounding datum.

The sounding datum is the datum to which depths on the charts and tide heights are referred, and is also known as the chart datum. It is reckoned through long-term tidal observations and chosen data models in designated sea waters, with the local

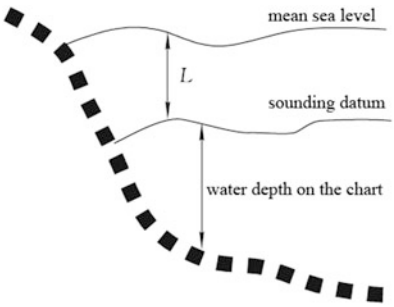


Fig. 3.15 The sounding or chart datum

MSL connected to the unified national height system. The location of the chart datum is typically selected at a vertical distance L below the local multiyear MSL (Fig. 3.15). The sounding datum should be determined by taking both the navigational safety and the rate of waterway utilization into account. Hence, it should be situated below the MSL and is approximately the tidal surface of the lowest water level. Countries worldwide have applied different formulae to calculate L based on the tidal regimes of the various waters, and there are hence a variety of chart datum choices, such as the theoretical lowest tide (TLT), lowest astronomical tide (LAT), mean lower low water (MLLW), lowest low water (LLW), mean lower low water springs (MLLWS), Indian spring low water (ISLW), mean low water (MLW), mean low water springs (MLWS), equatorial springs low water (ESLW), etc.

The tidal data of major ports are indicated on the chart, so it is possible to calculate the depth of water at a given point and a given time by adding the charted depth to the height of the tide at a particular moment found in the tide tables. When the chart datum is not compatible with the tidal datum, corrections also need to be applied.

The Sounding Datum Adopted by China

Prior to 1956, China adopted the lower low water as the chart datum. After 1956, the theoretically lowest tide level determined by Vladimirsky (abbreviated as the theoretical lowest tide) was taken as the chart datum.

3.2.3 Plans for Establishing China's National Vertical Control Network and Its Precision

The principles of establishing the national horizontal and vertical control networks are analogous. Vertical control networks are established based on four orders from higher to lower and from global to local, following the methods of stepwise control and densification. The leveling lines of different orders are generally required to be

in the form of closed loops, or close to leveling lines of a higher order in the form of loops to curb the accumulation of systematic errors and facilitate densification of the leveling lines of a lower order of accuracy.

The first-order leveling network is the backbone of the national vertical control network and the primary basis for relevant scientific research. Hence, the first-order leveling lines should be established along the traffic routes with gentle slopes in geologically stable areas in order to satisfy the needs of high-precision leveling. The leveling lines should be in the form of loops, and the loop circumferences in plain and hilly areas are of 1,000–1,500 km and about 2,000 km in mountainous areas. This density fits reasonably well for a country with as vast a territory as China.

The second-order leveling network is the overall basis of the national vertical control network. It should be established along railroads, highways, and rivers and should form loops. The loop circumferences are generally specified to be 500–750 km. In flat terrain, the loop circumferences can be decreased according to the needs of construction and increased in mountainous regions or harsh areas, taking into consideration the actual situations prevailing in each place.

Third- and fourth-order level networks provide the necessary vertical control points directly for topographic mapping and engineering construction. The third-order leveling lines are closed loops or annexed leveling lines densified within the higher-order leveling network with a loop circumference designated to be no more than 300 km. The fourth-order leveling lines generally form the annexed leveling lines, which start from and finish on higher-level benchmarks. The lengths of the annexed leveling lines are assigned to be no more than 80 km.

The precision in leveling of different orders is represented by M_{Δ}, the random mean square error for 1-km leveling, and M_w, the total mean square error of altitude difference. Their tolerances are as given in Table 3.1.

By the end of 1984, the field operations of the national first-order leveling network, which covered the whole mainland and Hainan Island of China, had been completed. Figure 3.16 shows the establishment of this leveling network.

The China's national first-order leveling network has 100 leveling loops and 289 leveling lines in total. The total length of the leveling line is 93,360.8 km. The overall adjustment was completed at the end of 1986. The "National Height Datum 1985" was adopted and the Qingdao leveling origin was defined as the reference point for heights, belonging to the normal height system. Precision in actual measurement are as detailed in Table 3.2.

Overall, repeated leveling measurements of China's national leveling networks were carried out over the years 1991 to 1997, concerned primarily with the original first-order leveling network. Specific leveling routes and nodal points were readjusted except for the spur lines. The total length of the re-leveled lines was approximately 94,000 km with 99 loops and 273 leveling lines. The precision targets after adjustment were that the random mean square error of altitude difference for 1 km was $M_{\Delta} \leq \pm 0.45$ mm and the total mean square error of altitude difference for 1 km according to closing errors of 99 loops was $M_w \leq \pm 1.0$ mm.

Table 3.1 Precision (M_Δ, in millimeters) of China's national leveling

Leveling order	First order	Second order	Third order	Fourth order
Tolerance of M_Δ	≤0.45	≤1.0	≤3.0	5.0
Tolerance of M_w	≤1.0	≤2.0	≤6.0	10.0

Fig. 3.16 China's national first-order leveling network

Table 3.2 Precision of leveling lines

Actually measured precision (mm)	<±0.3	±0.3 to ±0.5	> ± 0.51
Number of leveling lines	2	285	2

3.2.4 Leveling Route Design, Benchmark Site Selection, and Monumentation

Design

Technical design signifies to draw up work plans for establishing the most reasonably perfect leveling networks or leveling routes on small-scale maps according to the task requirements and conditions of the survey area. Thus, knowing the survey area thoroughly and collecting relevant data (like the topographic map of the survey area, previous leveling results, etc.) will be necessary before design starts. The leveling lines should be established along the traffic routes with gentle slopes to which benchmarks can be conveniently leveled. To decrease the chances of being

accidently disturbed, the leveling lines should be well away from the cities, railway stations, or other built-up areas, and be free from terrain obstacles like rivers, lakes, or valleys. It should be noted that when the intended leveling routes coincide with the old leveling lines, the old benchmarks should be made the best use of, if possible. New benchmarks should otherwise be set and connected to the old benchmarks.

Site Selection

After drawing up designs, reconnaissance surveys should be carried out to select routes and identify the site of benchmarks. The purpose of reconnaissance and route selection is to enable the design to accord with actual situations and to determine the viable leveling routes and benchmark sites. Considerations for selecting the benchmark sites include stability, security, long-term preservation, and accessibility. The benchmark sites should not be selected in areas vulnerable to floods, in soft soils, low-lying regions, or in places subject to fault movements. Locations where a benchmark cannot be found or be conveniently leveled to should also be avoided.

Monumentation

Benchmarks and identifiers will be used to mark the benchmark sites determined by reconnaissance survey over a long time to meet the needs of leveling connection. Benchmarks can be sorted into three categories according to their purposes: fundamental benchmarks, ordinary benchmarks, and benchmarks in bedrock. Their uses and underlying characteristics will be discussed separately below.

The fundamental benchmark serves to secure precise leveling results over a long time, in support of connection to the newly established benchmarks or for checks and renewals of elevations of previously established benchmarks. Primary benchmarks are established on the first- and second-order leveling lines, placed at intervals of 20–30 km and at approximately 60-km intervals in harsh desert areas.

The ordinary benchmark serves to provide an elevation reference for topographic and other engineering surveys directly. They are established on leveling lines of various orders at intervals of 4–8 km in general.

The benchmark in bedrock is a permanent benchmark directly anchored to the outcrop and is the primary basis for studying vertical crustal and surface motions. In general cases, the benchmark spacing on the first-order leveling lines is specified at 500 km. The number of benchmarks in megacities or near seismic zones can be increased to meet the requirements of scientific research.

Figures 3.17 and 3.18 show the specific forms, specifications, and installations of the primary and ordinary benchmarks.

After the leveling route design, benchmark site selection, and monumentation, we can obtain the height of the benchmark on the Earth's surface by the method of leveling.

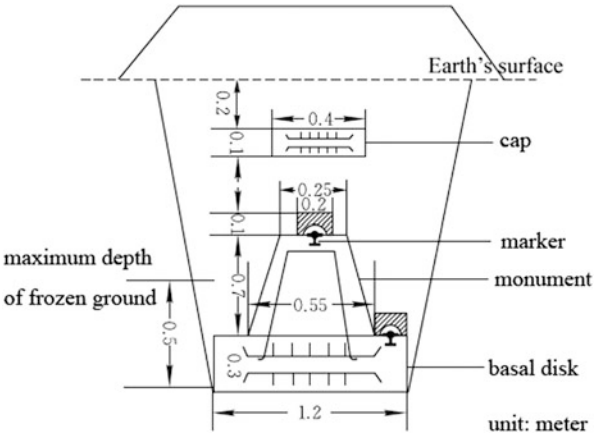


Fig. 3.17 Fundamental benchmark

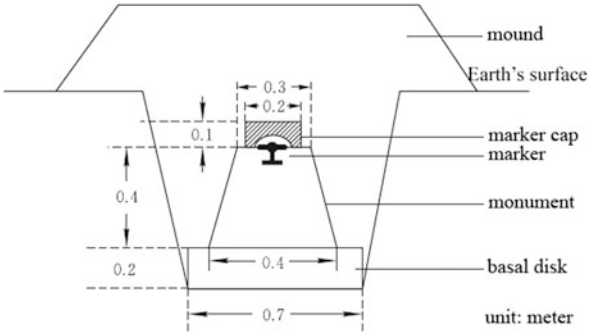


Fig. 3.18 Ordinary benchmark

Apart from leveling as the primary method for establishing vertical control networks, one can also employ trigonometric leveling and GPS measurement to obtain the height of a surface point. The latter two methods prove to be faster for establishing networks. The major flaw, however, lies in their lower accuracy compared to precise leveling.

3.3 The Three-Dimensional Coordinate Datum and Satellite Geodetic Control Networks

Similar to classical horizontal and vertical control networks, the satellite geodetic control network also consists of a series of geodetic control points on the Earth's surface in the form of a network. The difference lies in that through establishing a satellite geodetic control network, one can directly obtain the three-dimensional coordinates of control points, indicating realization of the three-dimensional geodetic datum. Therefore, the techniques used to establish control networks should be

space geodetic techniques, such as GNSS, VLBI, SLR, and DORIS, that can provide three-dimensional coordinates, rather than classical geodetic methods (i.e., triangulation, traversing, leveling, etc.), which cannot achieve such results.

3.3.1 The Three-Dimensional Coordinate Datum

In traditional geodetic surveys, horizontal and vertical coordinates are referred to different datums, which are inconsistent with each other, generating many problems in practical use. For example, research on geodynamic issues such as satellite and aircraft technology, the Earth's rotation, and plate movement must refer to the three-dimensional datum. Space geodetic techniques can realize the three-dimensional geodetic datum by determining or calculating the three-dimensional coordinates of surface points.

Global Three-Dimensional Coordinate Datum

The development of space geodetic techniques such as GNSS, VLBI, SLR, and DORIS has created conditions for establishing a global three-dimensional datum, which is realized by connecting the above-mentioned space geodetic stations distributed worldwide together into a corresponding global network. Each space technique is organized, coordinated, and managed by its corresponding international agency such as IGS, IVS (International VLBI Service for Geodesy and Astrometry), ILRS (International Laser Ranging Service), IDS (International DORIS Service), and so on. After the combined adjustment of the geodetic networks realized by each single space geodetic technique, a comprehensive three-dimensional datum, i.e., the International Terrestrial Reference Frame (ITRF) is formed.

Global IGS Network

In 1991, IAG decided to establish a global IGS observation network (originally known as the "International GPS Service" and changed into "International GNSS Service" in March, 2005). During the period from June to September, 1992, IAG implemented the first campaign for connection. Taking this opportunity, China held a nationwide "China 1992 GPS campaign" which was a cooperative endeavor involving many organizations. Major cities such as Shanghai, Wuhan, Taibei, Kunming, Xi'an, and Urumchi all participated in the IGS observation network. China's involvement in the global IGS cooperation aims to determine the high-precision geocentric coordinates nationwide and establish a new generation geocentric reference frame, as well as its transformation parameters with the national coordinate system. In addition, it also intends to determine the baseline vector

between stations with a relative accuracy of better than 10^{-8} order of magnitude and to establish the skeleton of a national high-precision satellite geodetic network. This work has laid the foundation for studies concerning crustal movement and geodynamics.

Being a global GNSS continuously operating station and an integrated service system, IGS provides users around the world for free with various GPS and Global Orbit Navigation Satellite System (GLONASS) information, such as satellite precise ephemeris, rapid ephemeris, broadcast ephemeris, velocities and coordinates of IGS stations, satellite signals received by IGS stations, and the Earth's rotation rate, etc. It has supported numerous scientific programs in geodesy and geodynamics, which include ionosphere, meteorology, reference framework, crustal movement, precision time-transfer, high-resolution calculation of the Earth's rotation rate and changes, and so on.

Base stations contributing to IGS are found all over the world and include more than 380 observation stations that belong to over 100 research institutions, universities, and governmental organizations in different countries. With the assistance of IAG, a coordination committee was set up by the agencies that have established the stations and the government departments of different countries, within which a board of directors was also formed. Authorized by IGS, this board of directors manages several Data and Analysis Centers and is in charge of providing services, launching products, and organizing international collaborative research projects. Its services and products include:

1. GPS satellite ephemeris service. Its products include broadcast, rapid, and final precise ephemerides with a precision of 25 cm, 5 cm, and better than 5 cm, respectively, and with a delay of 0 h, 17 h, and 13 days, respectively. GLONASS only provides the final ephemeris with a precision of about 30 cm.
2. Service of receiver clock error for GPS satellites and continuously operating reference systems (CORS). Its products are broadcast, rapid, and final clock errors with a precision of 5 ns, 0.2 ns, and 0.1 ns, respectively, and with a time delay of 0 h, 17 h, and 13 days, respectively.
3. Service of the coordinates (including the corresponding framework and epoch) and movement velocity of the continuously operating stations (tracking stations). Its products contain the horizontal and vertical positions and movement velocities of the station. The horizontal and vertical positions are at a precision of 3 mm and 6 mm, respectively, with corresponding yearly movement velocities of 2 mm per annum and 3 mm per annum.
4. The Earth rotation parameters service. Its products are rapid and final polar motion, short-term nutation, LOD (length of day) changes, etc. The accuracy of the final daily polar coordinates published by IGS is $\pm$ 0.1 milliarcseconds and the corresponding accuracy of the rapid product is $\pm$ 0.2 milliarcseconds.
5. Atmospheric parameters service. Its products are final tropospheric parameters, with the accuracy of zenith delay of 4 mm, issued delay $\leq$4 weeks. It will further provide the electron density distribution parameters of the ionospheric grid.

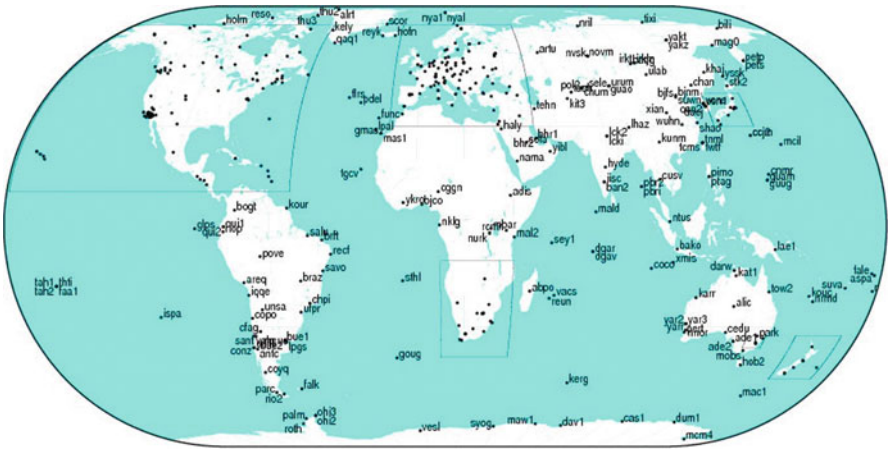


Fig. 3.19 Distribution of IGS stations. *Source*: IGS

The IGS network is the world's largest satellite geodetic control network, and its stations are distributed throughout the world as shown in Fig. 3.19. Apart from IGS, the IVS, ILRS, and IDS together have realized the International Terrestrial Reference Frame (ITRF) so as to realize the global three-dimensional geodetic datum.

Global ILRS Network

Presently there are more than 50 SLR stations all over the world and their distribution is shown in Fig. 3.20. To organize international connection and enhance cooperation, the International Laser Ranging System (ILRS) was established in November 1998 based on the original SLR Specialized Committee through reconsideration and reelection. Its central bureau was in GSFC (Goddard Space Flight Center) of NASA (National Aeronautics and Space Administration). The ILRS network is divided into four (several) subnets, which are introduced below.

NASA Network

In the late 1970s there appeared five stations such as MOBLAS (Mobile Laser Ranging System) stations, McDonald Observatory, Hawaii Haleakala station, etc. In the 1980s, four small mobile stations, TLRS 1–4, were added. Now, there are nine sets of equipment distributed in the continental USA, South Pacific (Tahiti), South America (Arequipa in Peru, TLRS 3), and Australia (Yarragadee). These stations, all equipped with the third-generation system, employ the active/passive mode-locked Nd:YAG (neodymium-doped yttrium aluminum garnet) laser produced by Quantel International (now known as Continuum). The pulse energy is 100–140 mJ (532 nm), the pulse width is 200 ps, and the repetition frequency is 4–5 Hz. As for the aperture of the telescope, MOBLAS is 76 cm and TLRS is

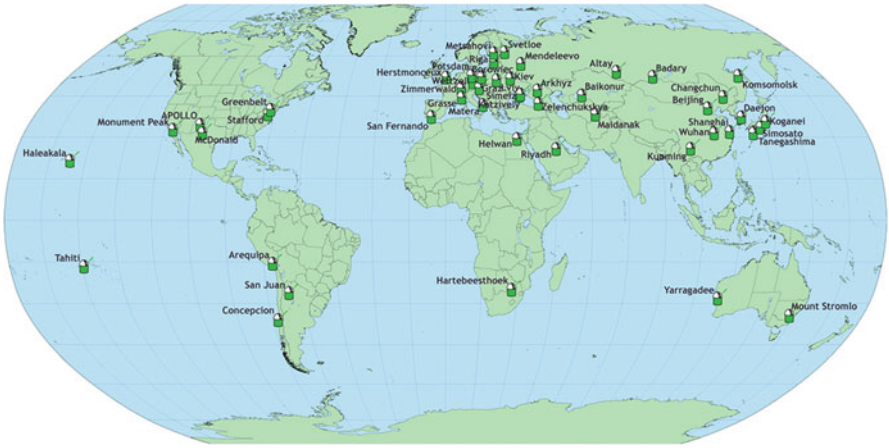


Fig. 3.20 Distribution of SLR stations around the world. *Source*: ILRS

25–30 cm. The electro-optic receiver utilizes a microchannel plate photomultiplier tube (MCP-PMT), receiving by multiphoton. The single-shot ranging precision with the LAGEOS satellites is 1–1.5 cm, of which MOBLAS has a high precision of 7–8 mm because of its strong echo, ranking top in the world. TLRS-2, with its small energy of 10 mJ and poor signal-to-noise ratio, has a less accurate single-shot ranging precision of 2 cm. Since the 1990s, small mobile stations have almost come to stagnation due to lack of funds. To make full use of the existing equipment, NASA redistributed five of the MOBLAS stations. In 1998, one was moved to Tahiti, an island in the South Pacific. Then in 2000, another one was moved to Hartebeesthoeck in South Africa.

The NASA network has retained its leading position in SLR throughout the world and has advanced technologies and high accurate ranging. Its number of observations represents about half of those of over 50 global stations that run normally.

EUROLAS

European Laser Consortium (EUROLAS), set up in 1989, now consists of 18 stations, of which the most important are Herstmonceux in Britain, Graz in Austria, Grasse in France, Wettzell and Potsdam in Germany, Zimmerwald in Switzerland, and Matera in Italy. Because the weather conditions in Europe are not as favorable as in America and Australia, the number of observations in European stations is comparatively smaller. However, the Herstmonceux station has a larger number of observations due to its better system stability. Equipment in the Graz station in Austria is very advanced, so the single-shot ranging accuracy of LAGEOS is 8 mm, ranking first in Europe. The weather conditions in the Grasse station are favorable, so the number of observations is accordingly greater. With a longer history and more advanced equipment, Wettzell station now possesses a set of large-scale

integrated measurement devices, Transportable Integrated Geodetic Observatory (TIGO), which includes various measurement methods such as SLR, VLBI, GPS, and PRARE, and instruments like gravimeters, seismometers, meteorological instruments, and so on. The SLR system in Wettzell uses the most advanced diode-laser-pumped Ti sapphire laser and is able to undertake dual-wavelength ranging. The Matera station recently installed a highly advanced SLR system with a telescope aperture of 1.5 m and the same ranging accuracy as that of NASA.

WPLTN

The West Pacific SLR Network (WPLTN) was set up in 1994. Its members include China, Japan, Australia, Russia, and Saudi Arabia, with 15 stations altogether.

There are two stations in Australia. One was the Orroral station of the Australian Surveying and Land Information Group (AUSLIG), located in Canberra and opened in the 1970s. Both its quantity and quality of observations ranked within the world's top five. It was closed in November 1998 and replaced by a new SLR system made by Electro Optic Systems (EOS) for AUSLIG. This new system was installed in Mount Stromlo, not far from the former station. It uses the advanced continuously diode-laser-pumped mode-locked oscillator from the Nd:YAG laser and a regenerative amplifier with pulse width of 25 ps. Their numbers of observations, ranging accuracy, and system stability currently all rank within the world's top three. It has been experimenting with unmanned automatic observation and has yielded initial results unprecedented in the world. The equipment in Yarragadee in the west of Australia is MOBLAS 5 of NASA, which has been turned over to AUSLIG. The weather conditions at this station are quite good so its number of observations is often among the world's top two.

The four SLR systems in Keystone, Japan, belong to the Communications Research Laboratory (CRL) and were made by EOS in Australia. They specialize in monitoring crustal deformation around Tokyo. The whole set of systems was highly advanced when it was put into use in 1996–1997. In 1999 it had realized a relatively stable ranging during the day with an accuracy of 1–1.5 cm. Unfortunately, these four stations no longer run normally due to lack of funds. The Simosato station of the Hydrographic Office of Japan is an old one that is still making regular observations.

Russia has more SLR stations, two of which are Komsomolsk and Maidanak (now belonging to Uzbekistan) and often participate in international cooperation. Their single ranging accuracy is about 4–6 cm. The number of observations per station per year is around 400–600 laps. In addition, two or three new stations have been successfully developed. With good performance, they will be used for regular observations.

The Chinese SLR Network comprises five fixed and two mobile stations (the equipment of some stations is shown in Fig. 3.21).

During 1971–1972, the North China Research Institute of Electro-optics (NCRIEO, in collaboration with Beijing Observatory) and Shanghai Observatory (working together with Shanghai Institute of Optics and Fine Machines) undertook

Shanghai Station

Beijing Station

Mobile Station

Fig. 3.21 Stations of the Chinese SLR network: Shanghai, Beijing, and a mobile station

SLR experiments for the first time in China. The first generation system adopted the Q-switched ruby laser, with a single ranging accuracy of 1–2 m. In 1980, for the first time Shanghai Observatory employed devices using Q-switched Nd:YAG lasers in the satellite ranging. The use of a constant fraction discriminator and a high precision event timer improved the ranging accuracy to 20–30 cm. In 1983, the second generation SLR system, which was organized by the Chinese Academy of Sciences (CAS) and completed under the coordination of several research institutions, was put into use in Shanghai Observatory. It detected the LAGEOS satellite 8,000 km away with a single-shot ranging accuracy of 15 cm. It also participated in the international Earth rotation connection survey of MERIT.

Changchun station of the CAS SLR network officially participated in international cooperation in 1992. In August 1997, the SLR system was improved considerably so that its single-shot ranging accuracy was improved from 5 cm to 1–2 cm and the quantity and quality of observations were both improved significantly. Its number of observations reached about 2,600 laps per year, ranking within the world's top ten. Beijing SLR station belongs to the State Bureau of Surveying and Mapping of China (SBSM) and has taken part in international cooperation since 1994. It has been upgraded since 1999 and now its ranging accuracy reaches 1–2 cm with observation data of 1,500 laps per year. Wuhan SLR station, jointly established by the CAS Institute of Geodesy and Geophysics (IGG) and Institute of Earthquake Science, China Earthquake Administration (CEA), began to join in

international cooperation in 1988. Because it was originally located in an urban area with harsh weather conditions, the observational data were modest. In 2000, the station was moved to the suburbs and thereby the conditions for observation were greatly improved. The CAS Yunnan Observatory began taking part in international cooperation in 1998. Now its ranging accuracy is about 3 cm. With a telescope aperture of 1.2 m, this system possesses strong ranging capacity and powerful laser energy and may possibly become a lunar ranging station in the future.

The two SLR mobile stations were both developed by the China Institute of Earthquake Science. One of them belongs to China Xi'an Research Institute of Surveying and Mapping and the other to the Institute of Earthquake Science. These two mobile stations are mainly used to monitor crustal movement in China.

The Chinese SLR Network was established in 1989 and was composed of the above-mentioned stations. It is now administrated by the CAS Shanghai Observatory, which is responsible for organizing and coordinating observations, unifying observation standards, and updating technologies in a collaborative manner. Shanghai Observatory is the SLR regional data center and data analysis center with the purpose of archiving domestic SLR data, making assessments of observations, and publishing a weekly assessment report on global observations. Meanwhile, it makes use of domestic and international SLR data to carry out research in the application of astro-geodynamics and geodetics.

Global IVS Network

VLBI observation is a kind of multistation network observation, which requires extensive international cooperation. To enable effective worldwide cooperation in VLBI observations and technological development, an international VLBI organization, IVS (International VLBI Service for Geodesy and Astrometry), was established. It serves to cooperate and organize the application of global VLBI in astrometry and geodynamics, enhance international cooperation in VLBI observations, data processing, and technological development, and provide services. IVS coordinates various kinds of activities in support of VLBI techniques. Its objectives are, first, to support research and observations in geodesy, geophysics, and astrometry; second, to promote studies and developments in VLBI techniques in astrometry and geodesy; and third, to allow communication between groups of users of various VLBI products and integrate VLBI techniques into the Global Earth Observation System. In China, both the Shanghai Sheshan station and the Urumqi Nanshan station are IVS stations. The IVS stations are distributed all over the world, as shown in Fig. 3.22. With regard to the different objectives of observation, the IVS Network can be divided into the following subnets and organizations:

European VLBI Network

This was the VLBI organization initially established by European countries. Since 1994, the Chinese VLBI stations in Shanghai and Urumqi have joined this

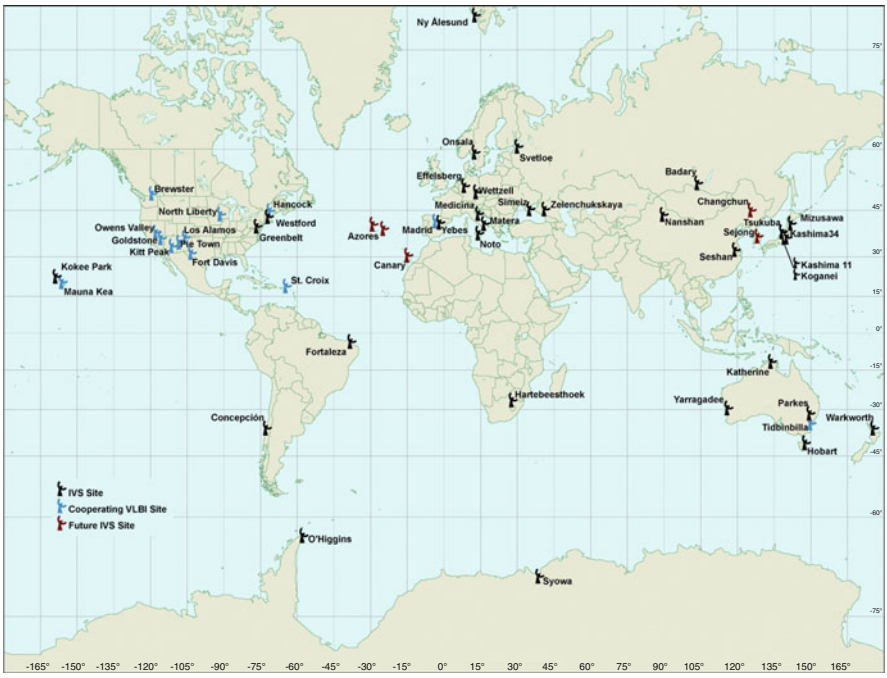


Fig. 3.22 Distribution of IVS network stations throughout the world. *Source*: IVS

organization. It has therefore become a Euro-Asian VLBI Network. European VLBI Network (EVN) provides observations on astrophysics and some astrometry projects as well as international cooperation in the development of VLBI techniques.

Asia-Pacific Telescope

This consists of VLBI organizations or observatories in the Asia-Pacific region and organizes astronomical and geodynamic VLBI observations as well as academic communications aperiodically every year.

Continuous Observation of the Rotation of Earth

This is a NASA project, managed by NASA's Goddard Space Flight Center (GSFC). Most VLBI stations with the capacity to carry out astrometry or geodetic surveying around the world have been involved in this project, whose scientific purposes are to measure continuously the Earth's rotation parameters with high precision and to provide highly precise data for the establishment and maintenance of the celestial and terrestrial reference systems, as well as the observation of modern plate movement.

VLBI Space Observatory Program

This is a space VLBI project held by the Institute of Space and Astronautical Science (ISAS) under the Japanese Ministry of Education, Science and Culture. It emits via an antenna of 8 m equivalent aperture to the Earth satellite orbit, forming a space VLBI station with an apogee of over 20,000 km. Most of the ground VLBI stations in the world have been involved in the space–ground VLBI observation of this project which has therefore become a global VLBI collaborative project.

VLBI Deep Space Exploration and China's VLBI Network

In recent years, as mankind's need for deep space exploration grows, there has appeared more concern over the role and significance of using VLBI to track and position man-made satellites and space objects. Europe and other countries like the USA and Japan have gradually established their own space VLBI networks and constantly update them. They have obtained satisfactory results in application by the VLBI network's tracking of space objects. The USA has successfully applied VLBI techniques to many space exploration projects (e.g., Apollo 16, Pioneer, Voyager, Galileo, etc.) due to its dominant position in space and its mature development regarding the application of VLBI techniques in space exploration. Its positioning accuracy has also been improved from 10 milliarcseconds at the beginning to better than 1 milliarcsecond in the 1990s. Japan's planned lunar exploration project (SELENE Project) will also adopt the VLBI technique and it has already carried out the relevant applied research.

China's VLBI began to develop in the late 1970s. In 1979, a proposal was officially raised on the establishment of China's VLBI and VLBI station system, including Shanghai station, Urumqi station, and Kunming station, and one relevant processing center. A 25-m radio antenna was also authorized to be built in Shanghai. The Sheshan station in Shanghai was completed and underwent test operation in November 1987. In April 1988, it began to carry out many international VLBI connection projects in astro-geodynamics, for instance, Sino-Japan collaborative observation, Sino-German cooperation program in VLBI geodesy, VLBI observations of the American (NASA) CDP (Crustal Dynamics Project), DOSE (Dynamics of Solid Earth), CORE (Continuous Observations of the Rotation of the Earth) programs, and APSG (Asia-Pacific Space Geodynamics) VLBI observations.

In October 1994, the Nanshan station in Urumqi was completed, allowing China's VLBI technique to be further developed. By the end of 1998, a VLBI mobile station with a 3-m aperture antenna, mainly set in Kunming, Yunnan, was put into test operation. The same year, the Shanghai Sheshan VLBI station was involved in the differential VLBI observations for positioning of the Mars Global Surveyor organized by NASA and yielded great results. In 1999, both the Sheshan and the Nanshan stations became the base stations of the CORE program. In 2003, these two VLBI stations succeeded in tracking and observing the launch process of

equatorial satellites in the Geospace Double Star Exploration Program (DSP) using the VLBI technique and obtaining high quality observational data.

In January 2004, the China National Space Administration declared that the Chinese Lunar Exploration Program (CLEP), or "Chang'e" project, was officially initiated, which meant that China's deep space exploration has entered an operational phase. In the Chang'e project, China's Unified S-Band Tracking Telemetering and Control (USB TT&C) system and CAS VLBI measurement system jointly monitored the orbits. Therefore, two VLBI fixed stations—one with a 40-m antenna in Kunming and the other with a 50-m antenna in Beijing—were formed on the basis of the Chinese former VLBI Network. Meanwhile, the functions of the former VLBI stations in Shanghai (with a 25-m antenna) and Urumqi (with a 25-m antenna) were transformed and extended. A rapid real-time data delivery channel between VLBI stations and related processing centers, real-time processors, and related processing centers were established as well. Now the Chinese VLBI Network has become a real-time observation network with four fixed stations and one relevant processing center (as shown in Fig. 3.23).

Due to cost-prohibitive equipment of the VLBI, SLR, and DORIS techniques, and the difficulties in setting observation stations, the stations realizing these techniques throughout the world are rather few in number and are mostly unevenly distributed. In consequence, it is difficult for them to be connected into an effective control network within certain areas. In practice they can only form a three-dimensional control network worldwide to realize the global three-dimensional datum by participating in the global connection survey through international cooperation. Comparatively speaking, the GNSS technique is characterized by convenient and rapid measurement and convenient station establishment. Therefore, GNSS is most often used in the establishment of regional three-dimensional control networks.

China's Three-Dimensional Coordinate Datum: High-Precision GPS Geodetic Control Network

For regional three-dimensional datum, one method is to establish a nationwide high-precision GPS network according to the needs of national construction and certain scientific research. Instances are China's high-precision A- and B-order GPS networks, the first- and second-order nationwide GPS networks, and the Crustal Movement Observation Network of China intended for earthquake prediction and plate motion research, etc.

A- and B-Order National GPS Networks

The A-order national GPS network was distributed by the SBSM, CEA and some other agencies in combination with the international IGS92 campaign in 1992. The entire network consists of 30 points distributed evenly in China's mainland with an

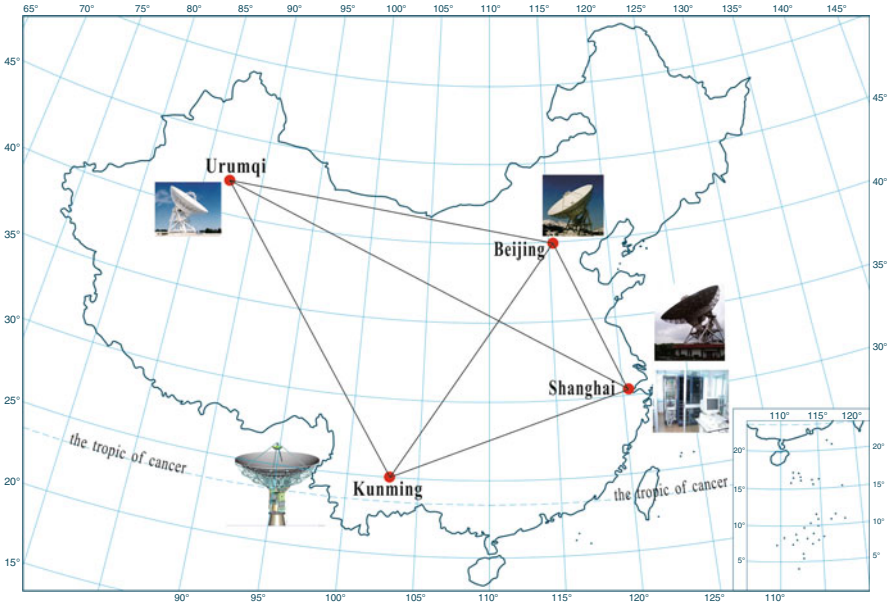


Fig. 3.23 Chinese VLBI network

average side length of about 650 km. In 1996, the repetition measurement of the A-order network was carried out by the SBSM. After overall adjustment of the entire network, the geocentric coordinate accuracy is better than 0.1 m, and the relative accuracy between points is better than 2×10^{-8} in the horizontal direction and better than 7×10^{-8} in the vertical direction.

The B-order network was distributed by the SBSM between 1991 and 1995. The number of points, together with those of the A-order network, adds up to 818. It is a continuous network with points of higher density in the east, a combination of a continuous network and a closed loop with points of moderate density in the center, and a closed loop and traverse lines with points of lower density in the west. Of all the points of the B-order network, 60 % coincide with China's first- and second-order benchmarks while the rest are under leveling connection. The repeatability accuracy between points of the B-order network is better than 4×10^{-7} in the horizontal direction and better than 8×10^{-7} in the vertical direction. The distribution of A- and B-order national GPS network stations is shown in Fig. 3.24. The coordinate framework adopted in the adjustment of A- and B-order networks is ITRF93 at the epoch of 1996.365.

First- and Second-Order Nationwide GPS Networks

To meet the requirements of military surveying and mapping and national construction, from 1991 to 1997, PLA Surveying and Mapping Bureau of the General

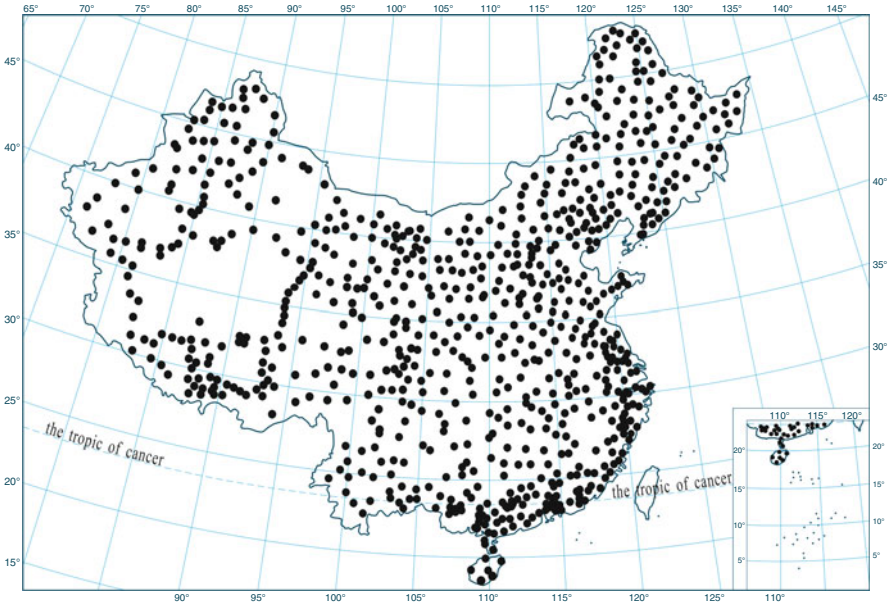


Fig. 3.24 The A-order and B-order national GPS networks

Staff Headquarters (SMBGSH) established high-precision GPS networks, including first- and second-order networks, which are called the first- and second-order nationwide GPS networks. Their scale and accuracy are about the same as that of A- and B-order national GPS networks. The distribution of their points is shown in Fig. 3.25. The complete networks consist of 553 points evenly distributed throughout China's mainland (except for Taiwan province) and in China's sea areas including the major reefs of Nansha Islands. The first-order network has 44 points with an average side length of 680 km and the observations were carried out between May 1991 and April 1992. The second-order network is divided into six survey areas (reefs in the South China Sea, Northeast, Northwest, and Southeast survey areas, North and East China survey areas, and the survey area of Qinghai-Tibet, Yunnan, Guizhou, and Sichuan) and the observations were carried out from 1992 to 1997. The second-order network is distributed on the basis of the first-order network with an average side length of 165 km. The points of both the first- and second-order networks are under leveling connection survey and are made to coincide with the national astro-geodetic network points. After adjustment computations, the accuracy of the first-order network is about 3×10^{-8}, and that of the second-order network is about 1×10^{-7}. The coordinate framework adopted in the adjustment of the A- and B-order networks is ITRF96 at the epoch of 1997.0.

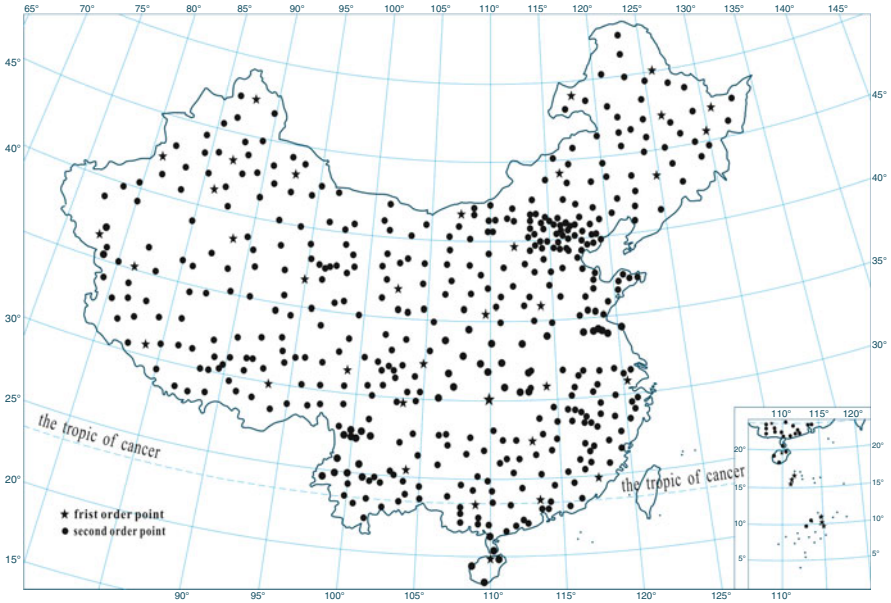


Fig. 3.25 First- and second-order national GPS networks

Crustal Movement Observation Network of China

The Crustal Movement Observation Network of China (CMONOC) was established from 1998 to 2002 by CEA, SMBGSH, and CAS (cf. Fig. 3.26). It is a monitoring network with earthquake prediction as its primary purpose. Its stations are mainly distributed on China's large plates and around its seismically active areas. CMONOC contains the datum network stations, the basic network stations, and the regional network stations, 1,222 stations in total. Among them, the average distance between datum stations (the continuous year-round observation GPS stations) is 700 km while that between basic stations (the periodic repetition observation stations) is 350 km. The datum and basic networks are mainly distributed on the larger plates in China. The regional network points are irregular repetition observation points, with spacing ranging from dozens to hundreds of kilometers. They are unevenly distributed across the country, more densely in areas with active crustal movement. The overview of CMONOC is shown in Table 3.3. The coordinate framework adopted in the adjustment is ITRF97 at the epoch of 2000.0.

China's National GPS Control Network 2000

Large-scale GPS networks, such as the established A- and B-order national GPS networks, the national first- and second-order nationwide GPS networks, and

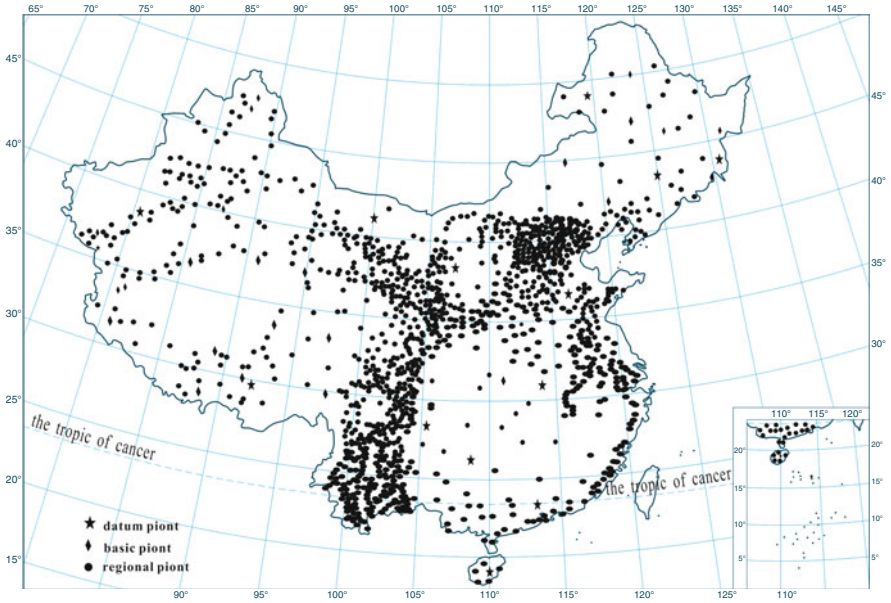


Fig. 3.26 Distribution of stations of the Crustal Movement Observation Network in China

Table 3.3 Overview of Crustal Movement Observation Network of China

	Datum network	Basic network	Regional network
Number of points	25	56	1,000
Distribution	Domestic plates	Domestic plates	Active areas of crustal movement
Observation	Continuous observation	Periodic repetition observation	Irregular repetition observation
Horizontal accuracy (mm)	–	2.5	1.8
Vertical accuracy (mm)	–	4.8	4.9
Accuracy of annual change in baselines (mm)	1.3	–	–
Precision of orbit determination (m)	0.5	–	–

CMONOC, have become the basic framework of China's modern geodetic survey and basic surveying and mapping, playing an increasingly prominent role in national economic development. The unique high precision of these networks helps improve and reinforce China's traditional astro-geodetic networks, thereby overcoming the disadvantages of the traditional astro-geodetic networks as having inhomogenous accuracy and possible larger systematic errors that traditional measurement methods cannot easily avoid. However, resulting from inconsistencies in the principles and purposes of network design, the observation guidelines, time of implementation, measuring instruments and standards, and methods of data

processing, there are differences in the datum and system between large-scale GPS networks. This has led to great difficulties in the practical application and made it difficult for GPS networks to work effectively as a whole.

To make full use of China's large-scale GPS networks, one must first eliminate the incompatibility of different networks and establish a unified systematic datum and a uniform, durable, and highly precise national GPS geodetic control network so as to establish a unified national geocentric coordinate system, improve accuracy of the geoid, and provide a better service for national economy, national defense construction, and geoscience studies. Hence, a unified adjustment to all the above-mentioned GPS networks was carried out by SMBGSH, SBSM, and CEA. After connection and data processing, the unified GPS network was named the "National GPS Control Network 2000" (Yang et al. 2009). The whole network has 2,609 stations, which can meet the requirements of modern measurement techniques for geocentric coordinates and lay a solid foundation for establishing China's new generation geocentric coordinate system (China Geodetic Coordinate System 2000) (see Chen et al. 2007a, b).

Continuously Operating Reference System

The Continuously Operating Reference System (CORS) is a major technique for establishing regional three-dimensional geodetic datums. Thanks to the rapid development of satellite positioning technology, information technology, and network technology, CORS networks have appeared all over the world, including global large-scale networks and regional small-scale networks. The classical horizontal datum is defined and extended by the geodetic origin and horizontal control network. Similarly, the three-dimensional coordinate datum is defined and extended by CORS and GPS networks.

Basic Components of CORS

As the datum of the three-dimensional control network, CORS is the product of sophisticated new technologies, such as satellite positioning technology, computer network technology, digital communication technology, and so on. It is composed of a datum station network, data processing center, data transmission system, positioning and navigation data broadcast system, and user application system. Every datum station is connected to the monitoring and analysis center through the data transmission system, constituting a special purpose network.

The datum station network, made up of datum stations distributed evenly within a certain scope, is used to collect observational data from GPS satellites, transmit them to the data processing center, and provide services for monitoring the system integrity.

The data processing center, as the control center of the system, is used to receive data from every datum station and form multiple datum stations' differential

positioning data for users through data processing. Then, the data files formed with a certain format are distributed to users. The data processing center is the essential part of CORS as well as the key to realizing high-precision real-time dynamic positioning. After a continuous around-the-clock solution of the whole modeling within an area based on real-time observational data collected by every datum station, the center automatically generates a virtual reference station (including the coordinates of datum stations and GPS observations) corresponding to the mobile station and provides various users who are in need of measurement and navigation with code phase/carrier phase differential correction information in a universal common format via the existing data communication networks and wireless data broadcast systems so as to calculate the exact real-time position of mobile stations.

The data transmission system, where data from every datum station are transmitted to the monitoring and analysis center through fiber lines. The system comprises hardware devices for data transmission and software control module.

The data broadcast system broadcasts positioning and navigation data to users in the form of mobile networks, UHF radio, Internet, etc.

The user application system consists of a user information receiving system, a network-based RTK (real-time, kinematic) positioning system, a fast and post-processing precision positioning system, an autonomous navigation system, positioning monitoring system, and so on. According to different application accuracies, the user service subsystem ranges from millimeter-level, centimeter-level, decimeter-level, to meter-level whereas, according to different user applications, it can be categorized into surveying and mapping and engineering users (centimeter- and decimeter-level), vehicle navigation and positioning users (meter-level), high-precision users (post-processing), and meteorology users.

CORS is not only a kinematic and continuous positioning reference frame but also an important technical means for rapid and high-precision acquisition of spatial data and geographic features. CORS is more often applied in small areas, for instance, a city, to provide three-dimensional datum. Within an area, it provides highly precise, reliable, and real-time positioning information to a large number of users simultaneously and unifies the surveying and mapping data in the city. For example, the application of CORS in cities will exert a profound and lasting influence on the collection of modern urban geographic information and the application system. Not only can it establish and maintain a reference frame of the surveying and mapping in cities, it can also provide automatically all-weather real-time spatial and temporal information of high-precision, serving as the basis of regional planning, management, and decision-making. It can also provide differential positioning information and develop new applications of transportation navigation; again it enables high-precision, high-spatial-and-temporal-resolution, all-weather, near-real-time, and continuous variable sequences in precipitable water vapor, which will gradually form a regional severe weather monitoring and forecasting system. CORS can also be used in the high-precision time synchronization in communication and power systems and will provide the services of monitoring and predicting land subsidence, geological disasters, and earthquakes and will study the temporal and spatial evolution of natural disasters.

CORS Networks in the World

Currently, CORS systems have been established or are being established in many developed countries worldwide. Here are some typical examples.

CORS Network in the USA

American CORS is in the charge of the National Geodetic Survey (NGS). It initially consisted of 137 GPS reference stations, such as the NGS tracking network, the United States Coast Guard (USCG) differential network, the WAAS (Wide Area Augmentation System) of the US Federal Aviation Administration (FAA), and the tracking network of the USA Corps of Engineers (USACE).

At present there are more than 2000 continuously operating GPS stations in the continental USA. They cover America completely (including Alaska) and constitute a new generation of American national kinematic reference system. All the reference stations in this system are equipped with full-wave double-frequency GPS receivers and ground rings antennae. Every day they unload the data of the day and record them in RINEX format of 1 s, 5 s, 15 s, and 30 s. This system offers reference station coordinates and data from GPS satellite tracking stations to users throughout the world, including Americans themselves, by means of the Internet. It also provides other services like geoid and coordinate system transformation. Users can observe any place in America with a GPS receiver, and can unload data from reference stations via the Internet to obtain post-processed precise positioning.

NGS can offer users corresponding GPS carrier phase and code distance (within the period of time users required) of CORS stations (more than three) adjacent to unknown GPS points through networks to support users' GPS near-real-time or post-processed positioning. Using networks, the NGS can also provide GPS positioning calculation services called Online Position User Service (OPUS), which will be completed within several hours of users offering the observational data of the unknown points.

Moreover, NGS networks can also provide its users with North American Datum 1983 (NAD83), the coordinates of ITRF and the displacement speed and meteorological data of its corresponding stations. NGS can also provide IGS precise and broadcast GPS ephemerides as well as the coordinates and displacement rate of IGS permanent continuously operating GPS stations all over the world. At present, the relative point position accuracies of the endpoints of the 26–300-km baseline calculated in America using CORS can reach 1.0 cm in the horizontal direction and 3.7 cm in the vertical direction, both with a 95 % confidence measure. The above case happens when users' observation at baseline endpoints lasts at least 4 h. If the observation time increases to 12 h, although the horizontal accuracy remains almost the same, the vertical relative accuracy can be improved to 2 cm.

Meanwhile, NGS also enables the transformation from NAD83 ellipsoidal heights to NAVD88 (the North American Vertical Datum of 1988) orthometric heights. It has primarily adopted the American geoid digital model to transform, and the error is ± 2.5 cm. In addition, GPS manufacturers usually only provide users

with the ARP (antenna reference point), which refers to the distance between the antenna pedestal and the antenna phase center when the GPS receiver receives L1 channel. Due to the difference in the GPS signal's direction of entrance and surrounding electronic field, the ARP positions will change with the electronic materials used by different GPS antennae manufacturers. Therefore, in order to improve the vertical positioning accuracy, NGS provides users with measured data concerning changes in antenna phase centers of various instruments for users' reference.

EUREF Permanent Network

Based on permanent satellite datum station networks established by European countries and some organizations (academic groups and universities), the EUREF Permanent Network (EPN) was established by the Regional Reference Frame Sub-Commission for Europe (EUREF) of the International Association of Geodesy (IAG) as a cooperative regional network of continuously operating stations. It is now composed of 122 permanent datum stations, among which 42 are IGS datum stations. Its working process is as follows: several permanent datum stations of different countries or organizations form subnets with their own operation centers; several operation centers constitute regional data centers whose data are gathered in the European regional center, which then transmits the product data back to IGS data centers, regional data centers, and various users.

At present, the assignment for EPN is to maintain the European Spatial Regional Reference Framework. The main service is to provide weekly computational results of the station coordinates throughout the network, the analysis of existing problems with their relative accuracy, and the analysis of time series weekly solutions. EPN also provides the original observed values of various stations to meet the requirements of precise positioning for different organizations in Europe. Apart from being the national spatial reference framework, CORS in European countries like Germany, Britain, and Switzerland still provides post-processed and real-time precise positioning based on differential and real-time kinematic (RTK) techniques.

COGRS System in the UK

The UK National Network of Continuously Operating GPS Reference Stations (COGRS) is similar to the above-mentioned CORS in terms of function and objectives. However, according to the situation in Britain, this COGRS system also monitors the relative and absolute changes in the sea level surrounding the British Isles. The UK COGRS is under the administration of its Ordnance Survey, Environment Agency, Meteorological Office, Ministry of Agriculture, and National Oceanography Centre. There are now nearly 60 GPS continuously operating active (reference) stations and over 900 GPS passive stations (points). At present, the UK Ordnance Survey has set up a GPS online service center, enabling transmission, supply, archiving, processing, and analysis of data from all GPS active and passive stations. This center also provides an online service of coordinate transformation.

SAPOS Network in Germany

The SAPOS (German Satellite Positioning Service) network consists of over 100 permanent GPS tracking stations. It provides four different levels of services: meter-level real-time differential GPS (precision, 1–3 m); centimeter-level real-time differential GPS (precision, 1–5 cm); near-real-time positioning (precision, 1 cm); and high-precision geodetic positioning (precision, better than 1 cm).

Other European countries with even a relatively small territory such as Finland and Switzerland have also established permanent GPS tracking networks with functions similar to the above-mentioned networks. As the datum for national geographic information systems, these tracking networks provide the scientific data for GPS differential positioning, navigation, geodynamics, and atmosphere.

COSMOS in Japan

COSMOS is the acronym for Continuously Operational Strain Monitoring System with GPS.

In Asia, Japan has established an integrated service system, namely the GPS Earth Observation Network System (GeoNet), which consists of almost 1,200 continuously operating GPS stations between which the average distance is about 30 km. This kind of grid, set up by the Japan Geographical Survey Institute, is an important national infrastructure for Japan. It mainly serves to carry out crustal monitoring and earthquake prediction, to form a high-precision kinematic national geodetic control network, to meet the needs of mapping and GIS data collection and updating, and to provide a service for departments of meteorology and atmosphere to carry out studies on GPS atmospherics.

CORS Networks in China

With the advancement of information technology, the rapid development of computer networks and communication technology, and the actualization of e-government, e-commerce, digital city, digital provinces, and digital Earth, there is a need for various real-time geospatial data to be collected. Therefore, it becomes more urgent and necessary for China to develop CORS networks. In recent years, different domestic industries have successively established some specialized continuously operating satellite positioning networks. Currently, to satisfy the needs for information for national economic development, a large number of cities, provinces, and industries are planning to establish similar continuously operating network systems. A boom in building CORS networks is coming.

The first CORS in China was established in Shenzhen (known as SZCORS) and has already started its all-round measurement. Some provinces and cities in China, such as Guangdong, Jiangsu, Beijing, Tianjin, Shanghai, Guangzhou, Dongguan, Chengdu, Wuhan, Kunming, and Chongqing, have initially established or are establishing similar provincial and municipal CORS .

At present, these high-precision GPS networks can be used as the basic geodetic control network for China. Datum stations of continuous operation in the Crustal Movement Observation Network can provide synchronous observation data for GPS measurements. With these stations as the initial points, GPS measurements will be more conveniently and efficiently carried out.

3.3.2 Establishment of Satellite Geodetic Control Networks

A scientific technological design of three-dimensional control networks is required in order to obtain reliable observation results. Typically, three-dimensional control networks such as VLBI and SLR are established worldwide by international organizations based on need. Thus, the process of establishment is rather complicated and time consuming. The so-called satellite geodetic control network usually refers to the three-dimensional control network in engineering applications; to be specific, the GPS control network. This section continues with an overview of the principles and technical design for establishing GPS control networks.

Principles for Establishment of GPS Control Networks

Establishment Based on Hierarchical Orders

Setting GPS network into different orders is conductive to stage-wise establishment according to the immediate needs and long-term development of survey areas. Moreover, this principle enables the network structure to combine the long and short sides. Compared to the Short-Side GPS Control Network, the network established in such a way can reduce the accumulation of errors at its edge and allows data processing and results checking of GPS networks to be carried out easily in a piecemeal fashion.

For instance, we can first use GPS to establish a nationwide high-precision backbone control network with low density (A- and B-order networks or first- and second-order networks) and then further densify the network using GPS or conventional methods based on the survey areas needed. In further densification, with the help of GPS technology it is unnecessary to establish an overall geodetic network in advance. Instead, one can establish and use the network at any time according to the accuracy required by users. We can obtain directly the known points from hundreds of kilometers away by GPS measurement, which not only saves a lot of manpower and material resources but also fulfills the practical needs.

Density

Different task requirements and service targets have different requirements for establishing the GPS network. For example, the national super-network (AA-order) datum points are mainly used to provide national datums for orbit determination, precise ephemeris calculations, and large-scale ground deformation monitoring, with an average distance of hundreds of kilometers. The network required by a general engineering survey with an average side length of several kilometers or even shorter (within hundreds of meters) should cater for the needs of mapping densification and engineering survey. Taking the above factors into account, a rule for the distance between two adjacent points in GPS networks is made dependent on various needs: the average distance between adjacent points in GPS at all orders should meet the requirements of the data in Table 3.4; the shortest distance between adjacent points can be 1/3 to 1/2 of the average distance while the longest is 2–3 times. Under special circumstances, depending on the network's task and target, the distance between some points can require specific rules for the distribution of GPS stations.

Accuracy

In the design of GPS networks, the order and accuracy standard should be designed based on the size of survey areas and the use of the networks. The accuracy standard of general GPS measurement is commonly expressed by the mean square error of the distance between adjacent points in the networks as follows:

$$\sigma = \pm\sqrt{a^2 + (b \cdot d)^2}, \tag{3.2}$$

where σ is the mean square error of distance (mm), a is the constant error (mm), b is the coefficient of the ratio error, and d is the distance between adjacent points (km).

The national "GPS survey specifications" classify GPS measurement into six orders, namely AA, A, B, C, D, E (as shown in Table 3.4). The Table lists the distances between points and their accuracy indicators in GPS networks of different orders (SBSM 2009).

Technical Design of GPS Control Networks

Design of GPS Control Network Datum

The design of the GPS control network datum is fundamental to the implementation of GPS measurement. It aims to find the best possible solution in terms of accuracy, reliability, and economic efficiency of the network. With GPS measurement we can obtain the GPS baseline vector between surface points, which belongs to the three-

Table 3.4 Accuracy and density of GPS control networks

	Order					
Item	AA	A	B	C	D	E
Constant error a (mm)	≤ 3	≤ 5	≤ 8	≤ 10	≤ 10	≤ 10
Ratio error coefficient b (ppm)	≤ 0.01	≤ 0.1	≤ 1	≤ 5	≤ 10	≤ 20
Average distance between adjacent points (km)	1,000	300	70	15 ~ 10	10 ~ 5	5 ~ 0.2

dimensional coordinate system of WGS84 or ITRF. Practical engineering applications require national coordinate systems like Beijing Coordinate System 1954, Xi'an Coordinate System 1980, China Geodetic Coordinate System 2000, or another independent local coordinate system. Therefore, in the technical design, the coordinate system and the initial data of the GPS network have to be specified, which means making clear the datum adopted by the GPS network.

The GPS network datum consists of position datum, azimuth datum, and scale datum. Position datum is usually determined by the coordinates of known initial points. Azimuth datum can be determined by the value of the known starting azimuth or the azimuth of the GPS baseline vector. Scale datum can be determined by the side of the electromagnetic wave distance measurement on the Earth's surface, by the distance between two initial points, or by the distance of the GPS baseline vectors. So, the design of the GPS network datum is essentially the issue of determining the position datum of the network.

Point Selection

Since GPS observation stations do not require intervisibility with each other, the selection of points is much simpler than for conventional measurements. The choice of GPS points has a significant influence on the smooth operation of GPS observations and the acquisition of reliable results. As a result, we should collect and fully understand the geographical conditions of the survey areas and the distribution and maintenance of existing control points based on the purpose of measurement and the requirements of coverage, accuracy, and density of the survey areas so as to properly choose the positions of the GPS points. The following principles should be followed in the selection of GPS point positions:

1. It should be convenient to install antennae and GPS receivers around the point. The point should be located where the view is not obstructed and the elevation angle of the surrounding obstacles is less than 15°.
2. The point should be far away from high-power radio emission sources and high-voltage wires to avoid interference from magnetic fields close to the signal.
3. In order to weaken multipath effects, there should be no objects that strongly reflect or absorb electromagnetic waves around the point.
4. To improve operational efficiency, the point should be located where transportation is convenient.

5. Points should be selected taking into account the convenience of using other measurement techniques for connection and extension.
6. The point should be located in solid soil or, better, an outcrop of rock in order to be better preserved.
7. The integrity and stability of the survey mark should be checked before using old points.

Additionally, other conditions such as the nearby communication facilities and power supplies should also be considered for the connections between points and the electricity for equipment.

Marking the Position of the GPS Control Point

For long-term preservation, the GPS control point should usually be located on the survey mark (monument) with an identifier in the center to mark the point precisely. Both the survey mark and the identifier should be stable and firm. The mark can be sunk into the ground or built into an observation stake or a stake with forced centering devices. For the structure, type, and methods of construction of survey marks, please refer to Sect. 3.1 and related technical standards.

Measurement Operations of GPS Control Networks

GPS measurement includes field observations and indoor data processing work. The former consists of installation of antennae, observation operations, and recording of field observation results whereas the latter consists of data extraction from instruments, baseline solution, and adjustment calculations of three-dimensional GPS control networks, etc. Below is an overview of field observations. For the extraction of observation data and the indoor work of data processing, please refer to annexed instructions and other relevant data.

Installation of Antennae

The precise installation of antennae is one of the prerequisites for precise positioning and should satisfy the following conditions:

1. Normally, the antenna should be installed in the vertical direction of the mark center on a tripod, directly centered. Only in exceptional cases can eccentric observations be allowed, where the centering elements should be precisely determined by analytical methods.
2. When installing an antenna on the platform of the tower, in order to avoid interference to signals, the top of the tower should be removed and the mark center be projected onto the platform. Then the antenna can be placed according to the projection point.

3. When there is an ordinary tower at the point and the distance between observation stations is less than 10 km, an antenna is allowed to be installed upon the tower, but the time of observation should be extended.
4. The pointer of the antenna should be directed to the true north. The effect of local magnetic declination should be considered and the orientation error should be no more than $\pm 5^{\circ}$.
5. The level bubble at the bottom of the antenna must be centered.
6. While installing an antenna in thunderstorm weather, the bottom of the antenna must be grounded to avoid lightning strikes. In a thunderstorm, observation operations should be stopped and the antenna removed.

After installation of the antenna, its height should be measured both before and after each time interval of observation. The difference between the results of the two measurements should be less than 3 mm. The average can be determined as the final height of the antenna. If the difference is larger than the tolerance, we should identify the causes, put forward suggestions, and note down the observation records.

The antenna height refers to the height from the average antenna phase center to the surface of the central mark of the observation station, which can be divided into two parts. One part covers the height from the phase center to the bottom of the antenna (i.e., the antenna reference point, ARP), which is a constant provided by manufacturers; the other part covers the height from the ARP to the surface of the central mark of the observation station, which should be measured by users on the spot. The specific measurement methods can be categorized into direct measurement and slant range measurement, according to the methods and types of antenna installation. Please refer to the receiver user manual for details. The final value of the antenna height is the sum of the heights of the parts.

Observational Operations

Observational operations are mainly aimed at capturing, tracking, receiving, and processing GPS satellite signals to obtain the required data on positioning and observations.

The operation of GPS receivers is highly automated. Its specific methods and procedures of operation vary with the types and operating modes of receivers. Detailed information is included in the attached operation manuals. During operation, observers only need to follow the instructions in the operation manual. Generally, the following aspects should be noted:

1. Observers at each receiver should work within the regulated observation time to ensure the realization of simultaneous observation of the same group of satellites.
2. After correct connection of the external power supply, the cable and antenna are confirmed, and the power can be switched on; when the receiver is in the right preset state it can be started.

3. When the data on the receiver's panel display is normal, observers can begin the self-test and input control information for observation stations and intervals of observation time.
4. When the receiver begins to record data, the observer should use function keys and selection menus to check information on observation stations, number of received satellites, satellite catalog number, channel SNR (signal-to-noise ratio), phase measurement residuals, results and changes in real-time positioning, records of storage media, etc.
5. During the period of observation, the receiver should not be turned off and restarted. The antenna height and limits of the elevation angle of satellites should not be changed. Observers are supposed to prevent vibration and especially displacement of the receiving devices. The antenna or signals should not be touched or obstructed.
6. When all the operation projects are confirmed to have been completed as required, the station can be moved.
7. In long-distance GPS measurements at higher levels, meteorological elements should be measured as required.

Observational Records

Observational records are automatically formed by GPS receivers onto storage media, which include carrier phase observations, pseudo-range observations, corresponding GPS time, parameters of GPS satellite ephemeris, clock offset parameters, and initial information of observation stations such as name, catalogue number, time intervals, approximate coordinates, antenna height, and so on. The information on observation stations is generally first input by observers into the receivers or recorded manually in measurement handbooks.

3.4 The Gravity Datum and Gravity Control Networks

To obtain the value of gravity quickly, we usually adopt relative gravity measurement. As a consequence, there must be some reference points with known gravity values that belong to a unified system. The point with known absolute gravity value is referred to as the gravity base station (gravity reference station), and the gravity value at the base station is known as the gravity reference value. The gravity datum is derived from a network formed by a series of gravity base stations. Just like the three-dimensional coordinate control network and horizontal and vertical control networks, the establishment of the gravity control network is also a basic project of constructing geodetic datums. Establishing high-precision gravity networks plays a prominent role in determining and refining the gravity field of the Earth and the geoid. This section is dedicated to the introduction of gravity datum, gravity basic networks, and establishment of gravity networks in China.

3.4.1 The Gravity Datum

Nowadays, the gravity datum of a network still requires a relatively small number of gravity reference stations to define the absolute datum level as well as the scale of the gravity net. Stations with known absolute gravity values serve as the reference points for relative gravimetry when determining the gravity difference between two stations. Such a reference point is also known as the gravity origin. The reference points acknowledged by the international surveying organizations are called the international gravity datum. Countries will do their utmost to correlate with the international gravity datum while determining gravity so as to test the accuracy of gravity measurements and to ensure that the surveying results stay unified. Universally recognized gravity datums include the Vienna Gravity System 1900, Potsdam Gravity System 1909, International Gravity Standardization Net 1971 (IGSN71), and International Absolute Gravity Base Station Network 1987 (IAGBN). These four gravity datums are briefly described below.

In 1900, the conference of IAG was held in Paris, at which the Vienna Gravity System was adopted, i.e., using the gravity value observed at Vienna Observatory, Austria as the datum. Its value for g is $(981.290 \pm 0.01) \times 10^{-2}$ m/s^2. This value was determined by Oppolzer in 1884 using absolute gravity measurement based on the reversible pendulum.

In 1909, the conference of IAG was held in London, at which the Vienna Gravity System was superseded by the Potsdam Gravity System. The Vienna gravity datum was replaced by the gravity value in the Pendulum Hall of the German Potsdam Geodetic Institute. The value of g is $(981.274 \pm 0.003) \times 10^{-2}$ m/s^2, which had been determined by Kuhnen and Furtwangler in 1898–1906 based on reversible pendulum measurements. The Potsdam Gravity System was the most widely applied. Almost all countries worldwide adopted the Potsdam gravity datum, which has been in use for 60 years.

With advances in technology, new requirements for accuracy in gravity measurements have constantly been raised. Since 1930, some countries have begun to research and develop the absolute gravimeter and carry out absolute gravimetry measurements. The number of absolute gravity stations has been growing throughout the world. The new absolute gravity stations have been connected to the Potsdam gravity datum using the relative gravimeter and the results have indicated a relatively large systematic error in the Potsdam gravity value of approximately 12×10^{-5} m/s^2 to 16×10^{-5} m/s^2. In 1967, IAG decided to add to the the Potsdam gravity system a correction of -14×10^{-5} m/s^2. During 1968 and 1969 a new absolute gravity measurement was made using the reversible pendulum at the Potsdam gravity origin, whose precision was up to $\pm 0.3 \times 10^{-5}$ m/s^2, improved by one order of magnitude. The result of the observed quantity differed by -13.9×10^{-5} m/s^{2}from the original gravity value.

At the 15th General Assembly of the International Union of Geodesy and Geophysics (IUGG), Moscow, 1971, a resolution was passed to replace the Potsdam datum with the IGSN71 as the new international gravimetric datum.

IGSN71 is the worldwide gravity datum. It encompasses 1854 gravity stations, ten of which are absolute gravity stations determined by three types of absolute gravimeters. It has over 25,200 measurement points for the use of relative gravimeters, among which more than 1,200 points are for pendulum measurement, and the rest use gravimeters. After overall adjustment of the observational results, gravity values of 1,854 stations, scale factors of 96 gravimeters, and zero drift rates of 26 instruments (pendulum meters and gravimeters) are obtained. The gravity value precision of each point after adjustment is $\pm 0.1 \times 10^{-5}$ m/s^2. Every point can be referred to as the initial point for gravity measurement; thus the multistation datum has put an end to the era of single-station datum (computed from a single origin of gravity).

As high-precision timing and ranging technology progressed, in the 1970s some countries successfully developed instruments to determine absolute gravity by the use of free-falling bodies. The accuracy of gravity measurement has been considerably improved. Many countries have initiated the establishment of their own gravity control networks rather than use the datum of IGSN71 as the initial point for gravity determination, and thus this gravity standard network has lost its control in reality. The precision of several microgals (1 μGal $= 10^{-8}$ m/s^2) has played a significant role in studying changes in the global gravity field. In addition, all absolute gravimeters suffer from certain systematic errors. Therefore, it is still imperative that the absolute gravity datums worldwide be unified.

After research and preparation over some time since the establishment of IGSN71, the plan for establishing IAGBN was proposed in 1982. At the 18th General Assembly of the IUGG 1983, the establishment of the IAGBN was decided upon to replace the IGSN71 with a very precise set of gravity stations. The primary objective of IAGBN is for long-term monitoring of the temporal gravity variations. It also serves as the gravimetric datum and provides conditions for gravimeter calibration. IAG has established a special study group and set strict requirements for the selection of absolute gravity stations. These established stations will undergo repeated observation at fairly regular intervals of time. IAGBN consists of two categories of stations, A and B. Stations of category A are selected based on the site selection criteria and the work plans. There are 36 such stations including one in Antarctica. Stations of category B are established for historical reasons or for the dream fulfillment of some countries. At the 19th General Assemble of IUGG, a resolution was passed to propose the initiation of the establishment of IAGBN. The majority of stations have been observed one or more times, but various circumstances have prevented the full implementation of the program at present. There are now two IAGBN stations of category A in China, located in Beijing and Nanning. Table 3.5 provides an overview of international gravity datums.

Table 3.5 Overview of international gravity datums

Name	Time of initial measurements	Time of application	Precision ($\times 10^{-5}$ m/s^2)
Vienna Gravity System	1884	1900–1908	±10
Potsdam Gravity System	1898–1904	1909–1971	±3
International Gravity Standardization Net 1971	1950–1970	1971–1983	±0.1
International Absolute Gravity Base Station Network	Established since 1983	–	±0.01

3.4.2 Basic Gravimetric Networks in China

In the late nineteenth century, foreigners conducted gravity measurements by means of a resilient pendulum in Shanghai and the southwest of China. In the 1930s, gravity measurement was carried out also using a resilient pendulum by the Physics Research Institute, Beiping Research Academy. Thereafter, Shanghai Petroleum Bureau determined several gravity points by means of a gravimeter in the vicinity of Shanghai. Up until the founding of P.R. China, the number of gravity points determined was only about 200, with accuracy of 5×10^{-5} m/s^2 to 10×10^{-5} m/s^2. These points were distributed in very limited areas, and no basic gravimetric networks were established then. After the founding of P.R. China, three generations of basic gravimetric networks were established in succession, i.e., China Gravity Basic Network 1957, China Gravity Basic Network 1985, and the National Basic Gravity Network 2000 as the third-generation gravity network in China.

China Gravity Basic Network 1957

During 1956 and 1957, in order to meet the needs of height anomalies and vertical deflections in data processing of the national astro-geodetic control network, China and the former Soviet Union jointly established the first-generation basic gravity network in China. No absolute gravity measurement was conducted then. Gravity values of the reference stations were measured by aerial survey from Moscow via Irkutsk, Almaty, and Chita base stations. Nine relative gravimeters were tied to the Beijing Airport in the west suburb for observation. Prior to this, the airborne gravity surveying group of the former Soviet Union had been conducting tie point observations between Potsdam and Moscow. The gravity station at Beijing Airport in the west suburb is the first gravity origin in China. It belongs to the Potsdam Gravity System, with an accuracy of $\pm 0.51 \times 10^{-5}$ m/s^{2}relative to the Potsdam international gravity datum. Meanwhile, 21 gravimetric basic points and 82 first-order gravity stations were established throughout China. The accuracy of tie point observations at the gravimetric basic points was $\pm 0.15 \times 10^{-5}$ m/s^2, and the accuracy of the first-order station was $\pm 0.25 \times 10^{-5}$ m/s^2. After adjustment of

these stations, the China Gravity Basic Network 1957 (abbreviated as CGBN57) took its shape. The error of its base station relative to the Beijing gravity datum was no more than $\pm 0.32 \times 10^{-5}$ m/s^2, and the error of the first-order stations was no more than $\pm 0.40 \times 10^{-5}$ m/s^2. The datum of this network has been introduced from three base stations of the former Soviet Union and belongs to the Potsdam Gravity System.

Nearly three decades since the establishment of the CGBN57, the departments concerned have measured tens of thousands of gravity stations of different orders. These gravity points are crucially important in national economic construction and national defense building.

In the early 1970s, the Chinese Academy of Metrology successfully developed the free-fall absolute gravimeter and initiated the first absolute gravity determination in China. Connected to the Beijing gravity origin, the observation proved that the original value was 13.5×10^{-5} m/s^2 too high. Therefore, in operations, once the Potsdam Gravity System is adopted, the correction of -13.5×10^{-5} m/s^{2}should be applied with no exceptions. Some departments add -14.0×10^{-5} m/s^2 to the gravity value of the Potsdam Gravity System according to the resolution of relevant international organizations.

China Gravity Basic Network 1985

The major problem of CGBN57 lies in that it has no absolute gravity station (generally known as a reference point or datum). Passing through many places, the gravity system was tied to Potsdam. At that time, the accuracy of the relative gravimeter was rather low. The Potsdam Gravity System was superseded by IGSN71 but China did not adopt this new system. Therefore, establishing a second-generation national basic gravimetric network was indispensable.

In 1981, China and Italy jointly measured 11 absolute gravity stations in China involving the use of free-fall absolute gravimeters from the Italian Institute of Metrology in accordance with the Sino-Italian Agreement on Scientific and Technological Cooperation. Under the organization of the State Bureau of Surveying and Mapping (SBSM), and with the participation of the Department of Geology and Mineral Resources, Department of Petroleum Resources, China Seismological Bureau (CSB), State Bureau of Metrology, Surveying and Mapping Bureau of the General Staff Headquarters of the Chinese People's Liberation Army (SMBGSH), Institute of Geodesy and Geophysics affiliated to Chinese Academy of Sciences (CAS), and some other departments, a connection survey of the national gravimetric basic network was conducted between 1983 and 1984. The overall surveying fell into two stages, employed nine gravimeters of LCR-G model, and was conducted in accordance with the "National Gravimetric Basic Network Field Operation Regulations (for Trial Implementation)." In terms of a relative connection survey, there should be at least four results from two instruments for each survey line and the mean square error of the average value should be no greater than $\pm 15 \times 10^{-5}$ m/s^2. Meanwhile, six LCR-G model gravimeters were employed to conduct the

international connection survey in Beijing, Shanghai, Paris, Tokyo, Kyoto, and Hong Kong. This has made possible the interconnection between the China national basic gravity networks and the international absolute gravity base station, IGSN71 station, as well as the international gravity tie points of the Japan Pacific Rim.

Adjustment was made in 1985 by the Chinese Academy of Surveying and Mapping, affiliated to the SBSM. The known base stations were Beijing, Shanghai, Qingdao, Fuzhou, Nanning, and Kunming. Considering the uneven distribution of these six base stations, the largest range of gravity only accounting for 65 % of the gravity range of the whole network, another five gravity points were used as the known points—Paris, Tokyo A, Tokyo B, Kyoto, and Hong Kong. The number of known gravity points was 11 in total. During adjustment, the observational values involved the conversion of recorded data from instruments and the corrections of instrument height, solid Earth tide, and atmospheric pressure. It was conducted by means of indirect adjustment of unequal weight. The China Gravity Basic Network 1985 (CGBN85) consists of 6 datum points, 46 basic points, and 5 points derived from the basic points. It has not only improved the graphical structure and offered external precision criteria, but also made the CGBN85 and IGSN71 closely interconnected, and brought the gravity system of CGBN85 into the system of IGSN71.

After overall adjustment, the mean square error of weight unit of the CGBN85 was $\pm 15 \times 10^{-8}$ m/s^2. The mean square error of the gravity value of the point (internal consistency) was $\pm 8 \times 10^{-8}$ m/s^2 to $\pm 13 \times 10^{-8}$ m/s^2. After checking the external consistency, the gravity values were found to be subject to some systematic effects. Therefore, the accuracy of the gravity values of the CGBN85 has been estimated to be somewhere between 20×10^{-8} m/s^{2}and $\pm 30 \times 10^{-8}$ m/s^2.

CGBN85 is the second national gravity control network in China, including a basic network and a first-order network. Its gravity datum has been jointly defined by the observed values obtained using many domestic absolute gravimeters and the known international gravity system. It must be pointed out that some of the datum values of the international gravity points are the observed values of absolute gravity, some are from the IGSN71 system, and some are the tie-in observations of the Japan Pacific Rim. Now we can see with hindsight that the gravity datum of this jointly defined CGBN85 was not defined independently by the absolute gravity points in China. It can only be described as something combined or synthetic.

China Gravity Basic Network 2000

The accuracy of the CGBN85, as opposed to the CGBN57, was improved by one order of magnitude, eliminating the error of the Potsdam System and increasing the density of basic points. China's basic gravity control network has played a crucially important role in various fields like surveying and mapping, geology, seismology, petroleum, and national defense since its implementation over a decade ago. However, with the passage of time, the economy has seen rapid growth, often making the basic points of the CGBN85 awkward or impossible to use. A survey

indicated that over two thirds of the CGBN85 gravimetric basic points cannot be used. Meanwhile, subjected to the constraints of the facilities and technologies at that time, CGBN85 has suffered from low accuracy of observation because of the absolute gravity station, uneven distribution of point position, and unreasonable graphical structure. All have proved that CGBN85 can no longer bring into full play the role of the national gravity datum.

China has introduced the FG5 absolute gravimeter with precision up to 3–5 $\times$ 10^{-8} m/s^2, which demonstrates a higher level of accuracy at the reference point of the Crustal Movement Observation Network of China (CMONOC). It provides the technical means for China to establish independently the new generation gravity datums of a higher order of accuracy. In addition, concerning the changes in gravity datums worldwide, it was decided that an international absolute gravity base station network is to be established. CGNB85, however, still belongs to the IGSN71 gravity system. All the above have necessitated establishing a new generation national basic gravity network (i.e., the China Gravity Basic Network 2000, abbreviated as CGBN2000).

In 1998, initiated by the China SBSM, with the participation of the SMBGSH and CSB, the joint establishment of CGBN2000 was started. With almost 3 years of arduous effort, the CGBN2000 was eventually established successfully in 2002. This network consists of 147 points, of which 21 are datum points (absolute gravimetric points) and 126 are basic points (relative gravimetric points). There are also 112 derived points and 66 ground tie points in cities. The points are distributed as shown in Fig. 3.27.

The targets of precision of the CGBN2000 after adjustment are: the mean square error of a gravity point in the basic network is $\pm$7.35 $\times$ 10^{-8} m/s^2, where the mean square error of a datum point with absolute gravity measurement is $\pm$2.3 $\times$ 10^{-8} m/s^2; the mean square error of a basic point is $\pm$6.6 $\times$ 10^{-8} m/s^2 and that of a derived point is $\pm$8.7 $\times$ 10^{-8} m/s^2. The mean square error of the 64 gravity points at the national station for gravimeter calibration is $\pm$3.4 $\times$ 10^{-8} m/s^2 and the mean square error of the 66 gravity points in the CGBN85 and the CMONOC and other networks connected to the CGBN2000 is $\pm$9.5 $\times$ 10^{-8} m/s^2

CGBN2000 comprises the datum points, basic points, derived points, and long and short baselines that are connected to the existing CGBN85. The network configuration is reasonable, taking into full consideration the needs of national basic construction, national defense construction, as well as disaster prevention and mitigation. It has a full range of various and complete functions and is also rationally and scientifically designed. This network boasts high precision and covers a large territory with a reasonably large number of points. The point positions also attend to China's actual situation. The quota of the points is appropriate, and these points are fundamentally evenly distributed. This network has absorbed advanced technologies at home and abroad and has adopted modern operational methods. It employs a rigorous theoretical approach to data processing, highly developed techniques, reliable adjustment results, and fairly credible precision. Compared to the CGBN85, this network has achieved a quality leap and

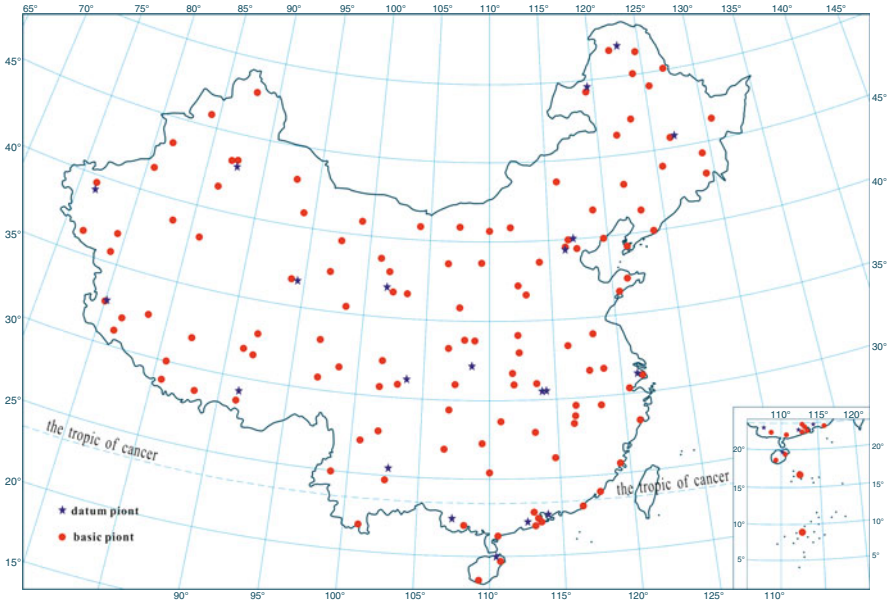


Fig. 3.27 Point distribution of China Gravity Basic Network 2000

reached a world-leading level. The next section gives an overview of the establishment of China's gravity network (see Table 3.6).

3.4.3 Establishment of China's National Gravity Networks

Absolute gravimetry primarily allows the determination of gravity for a small quantity of gravity base stations because the measuring equipment is bulky and costly. As a fundamental approach to gravimetric survey, relative gravity measurements are widely applied for the determination of gravity values on the Earth's surface. The discussion here is centered on the plans and methods for establishing gravity networks based on relative gravity measurements.

Fundamental Principles for Establishing China's National Gravity Networks

1. The national gravity network should cover all the provinces, autonomous regions, municipalities directly under the central government, the South China Sea, and Hong Kong and Macao Special Administrative Regions.
2. The absolute gravity stations in the net should be evenly distributed.

Table 3.6 Overview of China gravity networks

Name	Number of points			Measurement precision ($10^{-8}ms^{-2}$)		System
	Datum point	Basic point	I-order point	Datum point	Basic point	
CGBN57		21	82		±150	Potsdam system
CGBN85	6	46		±10	±20	IGSN71
CGBN2000	21	126		±5	±10	Absolute gravity system

3. Establishment of the gravity stations should give consideration not only to the needs of economic growth but also to the needs of national defense building and protection against and mitigation of earthquake disasters.
4. The network structure of the connection lines should be optimally designed.
5. The newly established gravity network stations should be connected to the old stations as well as the basic network of the crustal movement observation network of China (Fig. 3.26).

China has successively established the CGBN57, CGBN85, and CGBN2000 based on the above principles. Our discussions below will focus only on the plans for establishing regional gravity networks.

Plans for Establishing China National Gravity Networks

Gravimetry can be classified into two categories based on the purpose and accuracy of gravity measurements, i.e., "gravity control measurement" and "densifying gravity measurement." The former serves to establish control networks that include basic gravity points, first-order gravity points, and second-order gravity points. The latter, on the other hand, is conducted based on the gravity control point fulfilling the needs of the special tasks of various units and departments. Methods of connection survey and instruments used in gravity measurement vary with gravimetric order. This will be discussed further.

National gravity control networks should be established first before conducting gravity measurements in a country or region to obtain detailed gravity field data. A control network provides starting data for gravity measurements in areas of interest and helps with control of gravimetric error accumulation. As stated previously, China has successively established the CGBN57, CGBN85, and CGBN2000, which have offered the initial datum to perform nationwide relative gravity measurements. China has a vast territory, so the very few gravity control points of the CGBN2000 are apparently insufficient. It is hence necessary to extend further the first-order gravity network on the basis of the CGBN2000.

A first-order gravity point begins at a basic gravity point of known control in CGBN2000, the multi-line being connected pointwise to several unknown points

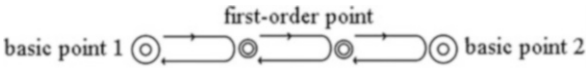


Fig. 3.28 Connecting line of gravity measurement

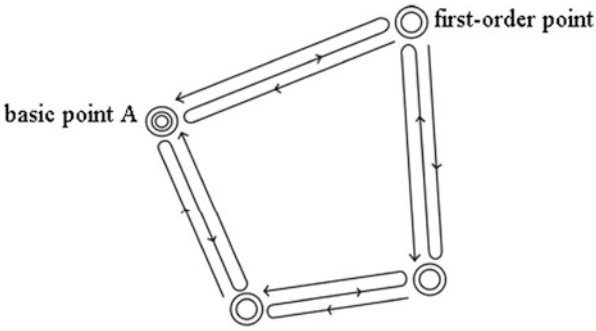


Fig. 3.29 Closed loop of gravity measurement

and terminating (closing) either at a second basic point, forming a connecting line (see Fig. 3.28), or at the same starting point, forming a closed loop (Fig. 3.29).

The point spacing of the first-order gravity network is approximately 300 km along the primary traffic routes. The number of line segments of the closed loop must not exceed five. The mean square error of the gravity difference between the line segments connected to each other must not exceed $\pm 25 \times 10^{-8}$ m/s^2. The first-order gravity points are required to be determined by LCR-G gravity meters or other precise gravimeters of equivalent quality. The constants and parameters of the instruments must be calibrated between the national basic gravity points or at the national-level stations for gravimeter calibration.

The second-order gravity point is a further extension of the basic network and the first-order gravity network. Its primary goal is to provide effective control for densifying gravity measurement. Hence, the establishment technique and density of the second-order gravity point can be determined according to the need for densifying gravity measurement. It is required that the higher-order gravity points and their derived points be used as initial points, established in the form of a closed loop or a connecting line. The line segments must generally not exceed five, but in harsh areas the number can reach eight. It is also acceptable to start from first-order gravity points and higher to develop second-order points in spur lines of one or two segments.

Gravity points in the densification network are established according to the different needs and complexity of the gravity field of the surveying area. The densification points are characterized by great density, small point-spacing, and fairly low-level accuracy. Therefore, gravimeters of any models installed at present can be adopted. The requirements can be readily satisfied using two instruments carried by automobiles to observe a survey line.

Determining the coordinates and heights of gravity points is an essential part of relative gravity measurements because the accuracy of coordinates and heights will directly affect the accuracy of gravity anomalies at gravity points. The mean square

error of height of the gravity point, as regulated, is not to exceed ± 2 m, whereas that of the coordinate is determined on a case-by-case basis. For the actually measured coordinate, its mean square error should not exceed ± 5 m. For the gravity points in the $1' \times 1'$ and $5' \times 5'$ grids, the mean square error is not to exceed ± 10 m, and not to exceed ± 100 m for the gravity control points and gravity points in grids greater than $5' \times 5'$.

If the gravity points coincide with geodetic control points and national benchmarks of different orders, then their coordinates and heights can be used directly. Otherwise, the coordinates and heights should be measured with GPS, traversing, leveling, or trigonometric height traversing.

For further and more exhaustive information on the specific implementation methods and process of gravity network establishment, please refer to the relevant literature on the subject.

Review and Study Questions

1. What is geodetic datum and how many datums does it consist of?
2. How many types of geodetic control networks are there? Briefly describe the methods for establishing each type of geodetic control network.
3. What are the principles of establishing the national horizontal control networks and the national leveling networks?
4. Into how many orders can the national horizontal control networks be classified? Outline the specifications (mean square error of angle observation, relative mean square error of side length, and average side length) for establishing the national horizontal control networks of different orders.
5. What are the specifications (constant error, ratio error) for establishing the national GPS networks of different orders?
6. Explain the concepts of zero elevation surface and leveling origin.
7. Describe the overview of development of the international gravity datum.
8. Briefly describe the plans for establishing the national gravity network.

Chapter 4
The Geoid and Different Height Systems

Theory of the Earth's shape (theory of the Earth's gravity field) provides the basis for determining the geodetic datums. The shape of the Earth can be defined in a number of ways. The true shape of the Earth is generally perceived as the natural surface of the Earth, i.e., the continental surface and idealized equilibrium sea and lake surfaces. Geodetic field operations are conducted on this surface. The mission of geodesy, however, does not involve acquisition of successive expressions of the Earth's surface, which should be the research domain of disciplines like cartography, aerial or space photogrammetry, and topography. The shape of the Earth in geodesy refers to the figure abstracted mathematically or physically from the true shape of the Earth, including the geoid, reference ellipsoid, and normal ellipsoid. The geoid can be understood as the physical shape of the Earth. A reference ellipsoid can be interpreted as the mathematical shape of the Earth, and a normal ellipsoid is the mathematical and physical shape of the Earth. The reference ellipsoid or normal ellipsoid is a close approximation of the geoid. Therefore, the shape of the Earth studied in geodesy primarily refers to the shape of the geoid. The geoid also serves as a reference surface for height determination of a given point on the Earth's surface. Since the geoid is the level (equipotential) surface of the Earth's gravity field, processing of leveling data should take into account the properties of the theory of the Earth's gravity field. The different vertical datums or reference lines chosen will constitute different height systems.

This chapter deals with the basic concepts of the theory of the Earth's gravity field, discusses the definition of height system, and establishes the relations of transformation between different height systems.

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_4,

4.1 Gravity Potential of the Earth and Geoid

4.1.1 Gravity and Gravity Potential

In light of Newton's law of universal gravitation, any two bodies in the universe, possessing mass, exert gravitational attraction on each other, which will thus create a gravitational field around the point mass. $\vec{F}$ is used to represent this attractive force. The force $\vec{F}$ is directly proportional to the product of their masses m and m' and inversely proportional to the square of the distance r between them, and can be expressed as:

$$\vec{F} = -\frac{Gmm'}{r^2}\frac{\vec{r}}{r}, \tag{4.1}$$

where G is the scale factor, referred to as the gravitational constant, which can be obtained through experiment. Its value is $6.67428 \times 10^{-11}\ \mathrm{m^3 kg^{-1} s^{-2}}$ (Pent and Luzum 2010). The direction of $\vec{r}$ is from the attracting mass toward the attracted mass (Fig. 4.1).

In geodesy, the particle of mass m is referred to as the attracting mass, while the other particle of mass m' is the attracted mass, the mass of which is used as a unit, i.e., $m' = 1$. Thus:

$$\vec{F} = -\frac{Gm}{r^2}\frac{\vec{r}}{r}. \tag{4.2}$$

The Earth can be regarded as a body constituted by infinite number of continuous point masses. The attraction that the Earth has exerted on the unit point mass $\vec{F}$ is the integral:

$$\vec{F} = -G\int_{Earth}\frac{1}{r^2}\frac{\vec{r}}{r}\mathrm{d}m, \tag{4.3}$$

where $\mathrm{d}m$ is the differential mass element of the Earth and $\vec{r}$ represents the position vector between $\mathrm{d}m$ and the attracted mass, which is a variable of integration; the integral area is the total mass of the Earth. The direction of the gravitational attraction is toward the center of the Earth.

Due to the rotation of the Earth, every point on the Earth experiences an inertial centrifugal force $\vec{P}$:

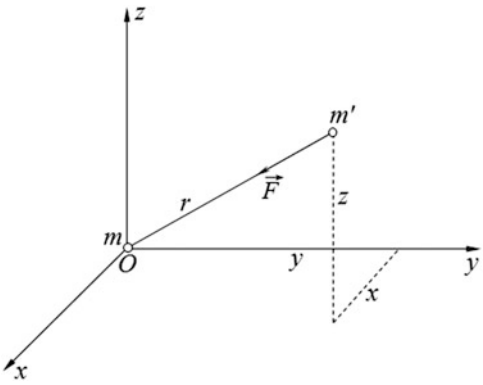


Fig. 4.1 The attracting mass m and the attracted mass m'

$$\vec{P} = m' \cdot \left(-\vec{\omega} \times \left(\vec{\omega} \times \vec{\rho}\right)\right), \tag{4.4}$$

where $\vec{\rho}$ denotes the vertical distance vector between the unit point mass and the spin axis of the Earth and $\vec{\omega}$ represents the angular velocity vector of the Earth rotation, which can be determined precisely using astronomical methods. Its value is $\omega = 7.292115 \times 10^{-5}$rad/s. $\vec{P}$ is perpendicular to the axis of rotation and is directed against the spin axis (Fig. 4.2).

The force of gravity of the Earth $\vec{g}$ is the resultant of the gravitational force acting upon a unit point mass and the centrifugal force of the Earth, namely:

$$\vec{g} = \vec{F} + \vec{P}. \tag{4.5}$$

Since the weight of a body is the product of its mass and the acceleration due to gravity, for one unit point mass the force of gravity acting upon it is equal to the value of its gravity acceleration. Therefore, in geodesy, the concepts of gravity force and gravity acceleration are always used interchangeably. When we say "to determine the force of gravity at a given point" we virtually mean to determine the gravity acceleration at this given point, and the magnitude of the force of gravity at a given point is actually the magnitude of its gravity acceleration. The gravity acceleration is measured in centimeters per second squared (cm/s^2), known as gal (after Galileo; symbol Gal) in geodesy. One thousandth of a gal is 1 mGal, and one thousandth of 1 mGal is 1 μGal, as follows:

$$1\,\text{Gal} = 1,000\,\text{mGal} = 1,000,000\,\mu\text{Gal},$$
$$1\,\text{mGal} = 10^{-5}\,\text{m/s}^2.$$

It is inconvenient to study the gravity vector directly. For any conservative gravity vector, there exists a so-called potential function such that the gradient of the function is the gravity vector. The partial derivatives of this function with

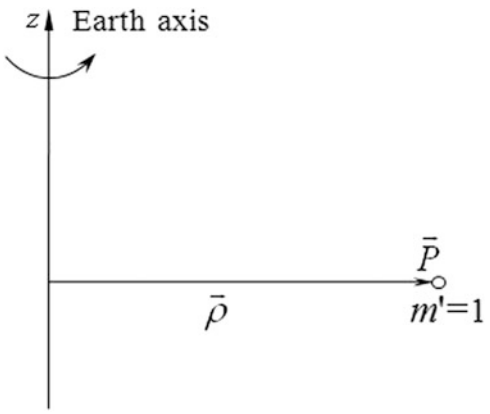


Fig. 4.2 Centrifugal force $\vec{P}$

respect to each coordinate axis are the gravity components along the coordinate axes. Therefore, we will use potential function in our discussion instead of gravity vector. Gravitational force, centrifugal force, and gravity force all have their corresponding potential functions.

The function of gravitational potential is a numeric function with respect to the variables of coordinate axes x, y, and z. Its partial derivatives with respect to the three coordinate axes correspond to the components F_x, F_y, F_z of the gravitational force $\vec{F}$ in these three directions respectively, namely:

$$\left.\begin{aligned} \frac{\partial V}{\partial x} &= F_x \\ \frac{\partial V}{\partial y} &= F_y \\ \frac{\partial V}{\partial z} &= F_z \end{aligned}\right\}. \tag{4.6}$$

The gravitational potential function at an exterior point of the body can be derived from a point mass gravitational potential function, which will be discussed first.

As shown in Fig. 4.1, m is the mass of the gravitationally attracting body at the point (0,0,0); m' is the attracted mass at the point (x, y, z); the distance r between them is given by:

$$r = \sqrt{x^2 + y^2 + z^2}.$$

Take a numeric function

$$V_{(x,y,z)} = \frac{Gm}{r}. \tag{4.7}$$

Apparently,

$$\left.\begin{aligned} \frac{\partial V}{\partial x} &= Gm\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = -\frac{Gm}{r^2}\frac{x}{r} \\ \frac{\partial V}{\partial y} &= Gm\frac{\partial}{\partial y}\left(\frac{1}{r}\right) = -\frac{Gm}{r^2}\frac{y}{r} \\ \frac{\partial V}{\partial z} &= Gm\frac{\partial}{\partial z}\left(\frac{1}{r}\right) = -\frac{Gm}{r^2}\frac{z}{r} \end{aligned}\right\}. \tag{4.8}$$

Comparing (4.8) with (4.2), it can be seen that (4.8) gives the components of the gravitational force $\vec{F}$ along the three coordinate axes. This indicates that the numeric function V in (4.8) is the gravitational potential function of a point mass.

It can be shown that the partial derivative of the potential function with respect to an arbitrary direction is the force component in the same direction. For instance, the partial derivative of (4.7) with respect to the direction r is:

$$\frac{\partial V}{\partial r} = -\frac{Gm}{r^2},$$

which has the length of the universal gravitational value and is similar to the universal gravity value.

In order to clarify further the physical meaning of V, in Fig. 4.1, assume that the unit point mass m' moves from point B_1 (distance r_1) to point B_2 (distance r_2); then the work (energy transfer) done by the gravitational force is:

$$A = \int_{B1}^{B2} -\frac{Gm}{r^2}\mathrm{d}r = \frac{Gm}{r}\Big]_{B1}^{B2} = \frac{Gm}{r_2} - \frac{Gm}{r_1},$$

where dr denotes the displacement in the direction of the force. The above equation indicates that the potential difference between the two points is the energy needed to move the point mass from the point of lower potential to that of higher potential. If the potential value at point B_1 is zero, then the potential of a point equals the energy needed to move the point mass from B_1 to this point.

Particle systems consist of a large number of point masses, and the gravitational potential is the sum of the gravitational potentials of the masses m_1, $m_2 \cdots m_n$ in (4.7):

$$V = \frac{Gm_1}{r_1} + \frac{Gm_2}{r_2} + \cdots + \frac{Gm_n}{r_n} = G\sum_{i=1}^{n} \frac{m_i}{r_i} \tag{4.9}$$

The mass is continuously distributed within the body, and hence it only requires conversion of the sum of (4.9) to an integral to obtain the formula for gravitational potential of the body:

$$V = G\int_M \frac{\mathrm{d}m}{r}, \tag{4.10}$$

where dm denotes the differential mass element at the point (ξ, η, ζ); it is a variable of an integral (see Fig. 4.3); $r = \sqrt{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2}$ is the distance from dm to the attracted mass and the total mass of the integral area is M.

The centrifugal force or acceleration is given by:

$$P = \omega^2 \rho, \tag{4.11}$$

where ω denotes the angular velocity of Earth rotation and ρ represents the vertical distance from the point being studied to the axis of rotation (see Fig. 4.2). Assume that the spin axis coincides with the z-axis of the rectangular coordinate system; then for the point (x, y, z):

$$\rho = \sqrt{x^2 + y^2}.$$

Inserting $\rho = \sqrt{x^2 + y^2}$ into (4.11):

$$P = \omega^2 \sqrt{(x^2 + y^2)}.$$

Obviously its potential function is:

$$Q = \frac{\omega^2}{2}\left(x^2 + y^2\right). \tag{4.12}$$

The force of gravity is the resultant of the gravitational force and the centrifugal force, and thus the gravity potential W is equal to the sum of gravitational potential V and the centrifugal potential Q, namely:

$$W = V + Q. \tag{4.13}$$

Hence the formula for gravity potential is expressed as:

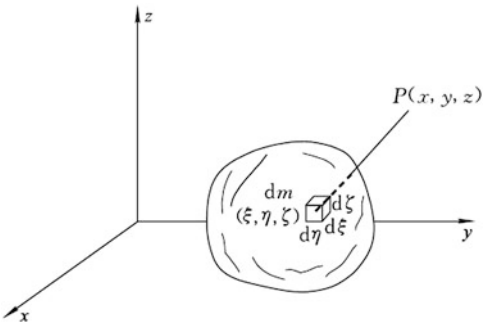


Fig. 4.3 Gravitational potential of the body with mass M

$$W = G\int_M \frac{\mathrm{d}m}{r} + \frac{\omega^2}{2}\left(x^2 + y^2\right). \tag{4.14}$$

4.1.2 Earth Gravity Field Model

It can be proved that the gravitational potential of a body at an exterior point, given by (4.10), that is:

$$V = G\int \frac{\mathrm{d}m}{r},$$

satisfies the differential equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0. \tag{4.15}$$

Equation (4.15) is the well-known Laplace equation. The function that satisfies the Laplace equation is termed the harmonic function (Torge and Müller 2012) or the spherical harmonics.

In the spherical coordinate system as illustrated in Fig. 4.4, the relation between the rectangular coordinates (x, y, z) and the spherical coordinates (ρ, θ, λ) of point P is given by:

$$\begin{cases} x = \rho \sin\theta \cos\lambda, \\ y = \rho \sin\theta \sin\lambda, \\ z = \rho \cos\theta, \end{cases}$$

where θ is the spherical colatitude. The Laplace equation (4.15) can be represented by the spherical variables (derivation omitted), as:

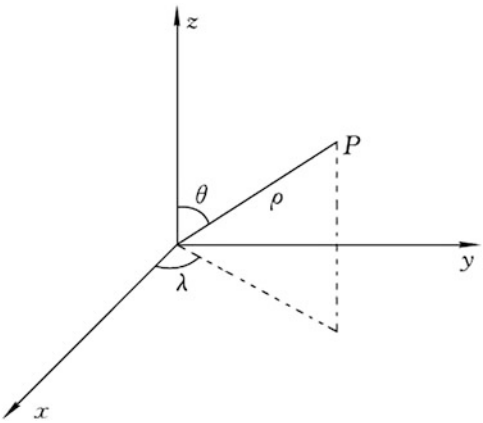


Fig. 4.4 Spherical coordinates and rectangular coordinates

$$\rho^2 \frac{\partial^2 V}{\partial \rho^2} + 2\rho \frac{\partial V}{\partial \rho} + \frac{\partial^2 V}{\partial \theta^2} + \cot\theta \frac{\partial V}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2 V}{\partial \lambda^2} = 0. \tag{4.16}$$

Solution of the above Laplace equation (derivation omitted) is the harmonic function in the spherical coordinates, which can be represented as (Hofmann and Moritz 2005; Torge 1989):

$$V(\rho, \theta, \lambda) = \sum_{n=0}^{\infty} \frac{1}{\rho^{n+1}} \sum_{k=0}^{n} (a_{nk} \cos k\lambda + b_{nk} \sin k\lambda) P_{nk}(\cos\theta). \tag{4.17}$$

Equation (4.17) is a series expansion, which indicates that the Earth's gravitational potential at an exterior point can be described by an infinite series. (ρ, θ, λ) are the spherical coordinates of the exterior point of the Earth and a_{nk} and b_{nk} are the coefficients of the Earth's gravity field, which can be determined by the observed values. Therefore, the gravitational potential problem can be regarded as the problem of study of the coefficient of the gravitational potential. $P_{nk}(\cos\theta)$ represents the associated Legendre polynomials (also known as the associated Legendre functions of the first kind), n is degree, and k is order. The expressions of the associated Legendre polynomials are given by:

$$
\begin{aligned}
&P_0(\cos\theta) = 1\\
&P_1(\cos\theta) = \cos\theta\\
&P_{11}(\cos\theta) = \sin\theta\\
&P_2(\cos\theta) = \frac{3}{4}\cos 2\theta + \frac{1}{4}\\
&P_{21}(\cos\theta) = 3\cos\theta\sin\theta\\
&P_{22}(\cos\theta) = -\frac{3}{2}\cos 2\theta + \frac{3}{2}\\
&\qquad\vdots
\end{aligned}
$$

The $P_{n0}(\cos\theta)$ in the above equation is simplified as $P_n(\cos\theta)$. $P_n(\cos\theta)$ is an nth-degree polynomial of $\cos\theta$, called the Legendre polynomial. The two polynomials $P_0(\cos\theta) = 1$ and $P_1(\cos\theta) = \cos\theta$ allow the higher degree polynomials to be generated using the recursion formula below:

$$
\begin{aligned}
&(n-k+1)P_{n+1,k}(\cos\theta) = (2n+1)\cos\theta P_{nk}(\cos\theta) - (n+k)P_{n-1,k}(\cos\theta)\\
&P_{nn}(\cos\theta) = (1-\cos^2\theta)^{\frac{1}{2}}(2n-1)P_{n-1,n-1}(\cos\theta).
\end{aligned}
$$

Coefficients a_{nk} and b_{nk} are related to the mass distribution and shape of the Earth, derived to obtain (Lu 1996):

$$
\begin{cases}
a_{n0} = G\displaystyle\int_{Earth} \rho_1^n P_n(\cos\theta_1)\mathrm{d}m,\\
a_{nk} = 2\dfrac{(n-k)!}{(n+k)!}G\displaystyle\int_{Earth} \rho_1^n P_{nk}(\cos\theta_1)\cos k\lambda_1\mathrm{d}m,\\
b_{nk} = 2\dfrac{(n-k)!}{(n+k)!}G\displaystyle\int_{Earth} \rho_1^n P_{nk}(\cos\theta_1)\sin k\lambda_1\mathrm{d}m,
\end{cases}
\tag{4.18}
$$

where $(\rho_1, \theta_1, \lambda_1)$ denote the coordinates of dm (Fig. 4.5). From the above equation, the meaning of the series expansion coefficients in the spherical harmonics of the potential of the Earth's gravitational field can obviously be further analyzed. It is commonly the first several terms of an infinite series that play a predominant role. The meaning of several coefficients in low-degree terms will be discussed below.

The zero-degree term has only one coefficient, a_{00}. Since $\rho_1^0 = 1$, $P_0(\cos\theta) = 1$, it follows from (4.18) that:

$$
a_{00} = GM,
$$

where M denotes the total mass of the Earth, which corresponds to the potential of the gravitational field generated by a homogeneous sphere of the Earth with its center at the coordinate origin.

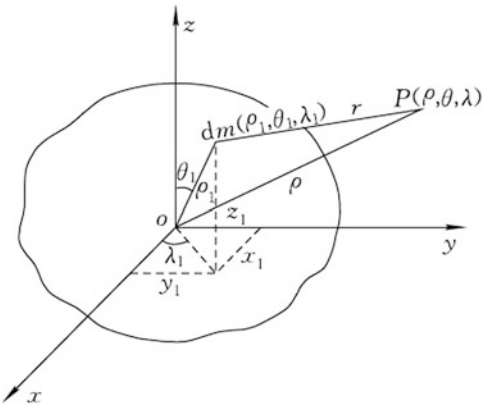


Fig. 4.5 Integral area is the entire Earth

The first-degree term has three coefficients, i.e., a_{10}, a_{11}, and b_{11}. For $P_1(\cos\theta_1) = \cos\theta_1$, and $P_{11}(\cos\theta_1) = \sin\theta_1$, it follows from (4.18) along with expressions of the spherical coordinates and the rectangular coordinates that:

$$
\begin{aligned}
a_{10} &= G\int_{Earth}\rho_1\cos\theta_1 \mathrm{d}m = G\int_{Earth} z_1 \mathrm{d}m,\\
a_{11} &= G\int_{Earth}\rho_1\sin\theta_1\cos\lambda_1 \mathrm{d}m = G\int_{Earth} x_1 \mathrm{d}m,\\
b_{11} &= G\int_{Earth}\rho_1\sin\theta_1\sin\lambda_1 \mathrm{d}m = G\int_{Earth} y_1 \mathrm{d}m.
\end{aligned}
$$

Assume the Cartesian coordinates of the Earth's center of mass are x_0, y_0, z_0; in the light of physics, its follows that:

$$
\frac{\int_{Earth} x_1 \mathrm{d}m}{M} = x_0, \quad \frac{\int_{Earth} y_1 \mathrm{d}m}{M} = y_0, \quad \frac{\int_{Earth} z_1 \mathrm{d}m}{M} = z_0
$$

Hence the three coefficients of the first-degree term are:

$$
a_{10} = fMz_0, \quad a_{11} = fMx_0, \quad b_{11} = fMy_0
$$

It is thus clear that the first-degree terms are related to the center-of-mass coordinates of the Earth. If we place the origin of the coordinate system at the Earth's center of mass, then the numerical values of these terms are zero.

The second-degree term has five coefficients, i.e., a_{20}, a_{21}, a_{22}, b_{21}, b_{22}. Integrating (4.18) gives:

$$a_{20} = G \cdot \left(\frac{A+B}{2} - C\right), \quad a_{22} = G \cdot \left(\frac{B-A}{4}\right),$$

where A, B, and C indicate the moments of inertia of the Earth with respect to the X-axis, Y-axis, and Z-axis, respectively, namely:

$$A = \int_{Earth} \left(y_1^2 + z_1^2\right) \mathrm{d}m, \quad B = \int_{Earth} \left(x_1^2 + z_1^2\right) \mathrm{d}m, \quad C = \int_{Earth} \left(x_1^2 + y_1^2\right) \mathrm{d}m.$$

The other three coefficients are:

$$a_{21} = G \int_{Earth} z_1 x_1 \mathrm{d}m, \quad b_{21} = G \int_{Earth} y_1 z_1 \mathrm{d}m, \quad b_{22} = \frac{1}{2} G \int_{Earth} x_1 y_1 \mathrm{d}m.$$

The above three integrals are the products of inertia about the X-axis, Y-axis, and Z-axis. Therefore, the second-degree term is dependent on the moments of inertia and products of inertia of the Earth with respect to the coordinate axes.

The coefficients of the terms of the third degree or higher are rather complex and are not discussed here.

Substituting the above nine coefficients into (4.17), putting the origin of the coordinate system at the Earth's center of mass, and making the coordinate axes coincide with the Earth's principal axis of inertia, the coefficients of the first-degree terms, a_{21}, b_{21}, and b_{22} in the second-degree terms are all zero, and then the expansion of the Earth's gravitational potential is:

$$V_{(\rho,\theta,\lambda)} = \frac{GM}{\rho} + \frac{G}{\rho^3}\left[\left(\frac{A+B}{2} - C\right)\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) + \frac{3(B-A)}{4}\cos 2\lambda \sin^2\theta\right] + \\ + \sum_{n=3}^{\infty} \frac{1}{\rho^{n+1}} \sum_{k=0}^{n} (a_{nk}\cos k\lambda + b_{nk}\sin k\lambda) P_{nk}(\cos\theta) \ .$$

In practice, the spherical harmonics expansion of the gravitational potential of the Earth is written as:

$$V_{(p,\theta,\lambda)} = \frac{GM}{\rho}\left[1 - \sum_{n=2}^{\infty}\left(\frac{a}{\rho}\right)^n J_n P_n(\cos\theta) + \sum_{n=2}^{\infty}\sum_{k=1}^{n}\left(\frac{a}{\rho}\right)^n \right. \\ \left. \left(\overline{J}_{nk}\cos k\lambda + \overline{S}_{nk}\sin k\lambda\right)\overline{P}_{nk}(\cos\theta)\right], \tag{4.19}$$

where a denotes the semimajor axis of the Earth ellipsoid and $\overline{P}_{nk}(\cos\theta)$ is the fully normalized associated Legendre polynomials, which differ from the associated Legendre polynomials only by a constant factor:

$$\overline{P}_{nk}(\cos\theta) = \sqrt{2(2n+1)\frac{(n-k)!}{(n+k)!}}P_{nk}(\cos\theta) \quad (k>0)\,.$$

Since the values of the associated Legendre polynomials differ very much from each other when the degrees of the polynomials vary considerably, such as $P_{21}(\cos 58\ ^\circ) = 1.3482$, and $P_{88}(\cos 58\ ^\circ) = 542279$, the results of the high-degree polynomials computed using recursion formulae may lead to relatively large accumulative errors, whereas $\overline{P}_{21}(\cos 58^\circ) = 1.7405$, and $\overline{P}_{88}(\cos 58^\circ) = 0.6913$. $J_n, \overline{J}_{nk}, \overline{S}_{nk}$ are the normalized coefficients. Comparing (4.19) with (4.17), we can obtain:

$$J_n = -\frac{a_{n0}}{GMa^n}$$

$$\overline{J}_{nk} = \frac{a_{nk}}{GMa^n}\cdot\sqrt{\frac{(n+k)!}{2(2n+1)(n-k)!}}$$

$$\overline{S}_{nk} = \frac{b_{nk}}{GMa^n}\cdot\sqrt{\frac{(n+k)!}{2(2n+1)(n-k)!}}$$

Adopting this set of coefficients, the differences between the values of the polynomials changing along with n and k (degree and order) are slight, making it convenient to use. The coefficient that is independent of longitude in (4.19) is called the coefficient of zonal harmonics, while the coefficient that is longitude dependent is called the coefficient of tesseral harmonics.

As the potential of the gravity field W is the sum of the potential of the gravitational field V (4.19) and that of the centrifugal field Q (4.12), the mathematical expression of W is referred to as the Earth gravity field model.

4.1.3 Level Surface and the Geoid

We know that every point P on the Earth will be acted upon by the inertial centrifugal force $\vec{P}$ and the Earth's gravitational force $\vec{F}$ (Fig. 4.6). The resultant of these two forces $\vec{G}$ for a unit mass is called gravity. The action line of gravity is referred to as the plumb line. The direction of the force of gravity is the direction of the plumb line. Given the inhomogeneous distribution of matter within the Earth and the undulating form of the Earth's surface, the changes in the direction of the plumb line at each point are irregular and the plumb line is far from a straight line.

When a fluid is in a state of equilibrium, everywhere on its surface is normal to the direction of gravity, otherwise the liquid will flow. We call this the equilibrium surface of the liquid level surface. Gravity exists everywhere on the Earth and in

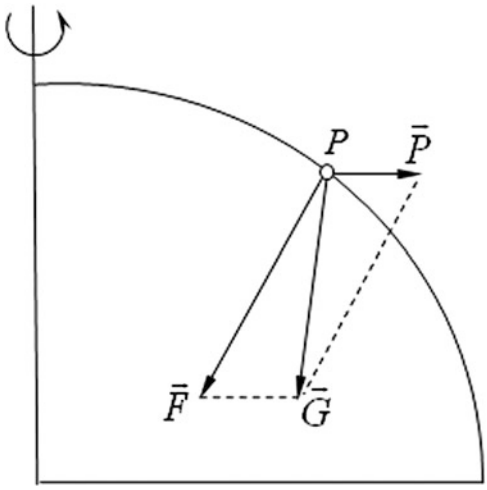


Fig. 4.6 Direction of the plumb line

space, so points at different heights will result in different level surfaces. The level surface is a physical reality because it is a surface of constant geopotential, i.e., no work is done in moving a mass along a frictionless level surface. The level surface is also known as the equipotential surface of gravity. The vertical at each point on the level surface is orthogonal to the plane.

When angles are measured using theodolites, or height differences are determined using leveling instruments, the instruments should all be leveled. While observing the horizontal angle, the bubble in the spirit level of the theodolite should rest in the middle. Thus, by centering the bubble, the line of sight is made horizontal. The vertical axis of the instrument coincides with the direction of the plumb line. The horizontal plane measured by the horizontal circle is a plane tangent to the level surface. Therefore, the horizontal angles actually observed are the angles on level surfaces at different heights. Likewise, the height difference determined using leveling methods is the distance along the plumb line between level surfaces. The astronomical longitude and latitude as well as astronomical azimuth also refer to the level surface and plumb line. Thus, the level surface and plumb line are the reference surface and the datum line for field operations using theodolites, leveling instruments, and such optical measuring instruments.

There are many equipotential surfaces. We define one of them as the geoid.

The geoid is the equipotential surface, which approximately coincides with the mean sea level (MSL) in the ocean and its extension under the continents. Oceans cover about 71 % of the Earth's surface, and the average elevation of continents is about 800 m above MSL, about one eight-thousandth of the Earth's radius, which means that the geoidal body closely approximates the natural surface of the Earth. As a result, it is natural to employ the geoid as a representation of the Earth. The geoid also serves as a reference surface for height determination of a given point on the Earth's surface while studying the shape of the Earth's surface. Meanwhile, the geoid is also employed as the reference surface for reduction of the astronomical longitude, latitude, azimuth, and the values of gravity.

The geoid is an irregular curved surface. The magnitude and direction of the gravitational force differ from point to point due to the terrain undulations and inhomogeneous distribution of mass inside the Earth, which will also cause irregular direction changes of the plumb line at different points. Hence, this geoid, everywhere perpendicular to the direction of the plumb line, is correspondingly an irregular curved surface with slight undulations. Therefore, the geoid is a physical rather than a mathematical surface.

With the in-depth studies of oceanography, people have realized that the MSL and the geoid are different concepts. The MSL is not the level (equipotential) surface, for many factors can exert influence on the oceans such as temperature and pressure variations, salinity, winds, currents, rotation of the Earth, etc. Meanwhile, the MSL measured using tide gauges in different countries or areas also varies. If a certain equipotential surface is chosen as the standard sea level, then separation between MSL and the standard sea level is referred to as the sea surface topography or sea surface slope. The rise and fall is about 1–2 m on a global scale. For the eastern sea area of China, the sea surface slope has high elevation in the south and low elevation in the north. The height difference is approximately 60 cm.

It is imprecise to define the MSL as the geoid, for the MSL is not an equipotential surface. Therefore it is recommended that the geoid be defined as an equipotential surface passing through the point to which heights are referred (zero level).

Based on the properties of potential function, we can conclude that the derivative of gravity potential W with respect to an arbitrary direction s is equal to the component g of gravity g_s in the same direction, namely:

$$\frac{\mathrm{d}W}{\mathrm{d}s} = g_s = g\cos(g, s) \tag{4.20}$$

In the case that the direction s is perpendicular to that of gravity, $\cos(g, s) = 0$, then:

$$\frac{\mathrm{d}W}{\mathrm{d}s} = 0.$$

The integration yields:

$$W_{(x,y,z)} = \text{constant}. \tag{4.21}$$

Assigning a fixed value to the constant on the right-hand side of the equation will yield an equation of the curved surface. It is referred to as the equipotential surface given that the gravity potential value is equal everywhere on this surface. In the meantime, the direction of gravity at any point intersects the surface. The surface is in an equilibrium state, i.e., the level surface (also the equipotential surface or constant geopotential surface) of gravity. The geoid is the level surface passing through the reference point for heights.

Some properties of the level surface can be further studied by applying the concept of gravity potential.

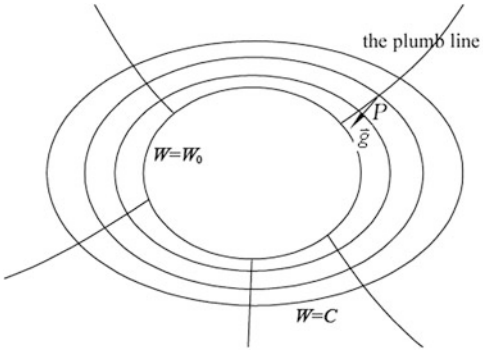


Fig. 4.7 Level surface

In (4.20), if s is directed against the direction of gravity g, denoted by h, in this case, $\cos(g, h) = -1$, yields:

$$\frac{\mathrm{d}W}{\mathrm{d}h} = -g,$$

also written as:

$$\mathrm{d}h = -\frac{\mathrm{d}W}{g}, \tag{4.22}$$

where dW represents the potential difference of two infinitely close level surfaces, and dh is the vertical distance between the two level surfaces. Equation (4.22) indicates that the distance between level surfaces is inversely proportional to gravity.

Because g varies everywhere on a level surface, the level surfaces are non-parallel. Figure 4.7 is a general description of the level surface. In the meantime, the value of g is finite and dh cannot possibly be zero, so the level surfaces do not intersect. However, within a small area the gravity value does not change much, and the two level surfaces can be considered parallel; for instance, the two level surfaces at each point, where the leveling rods are held, are considered parallel to each other in leveling. Hence, the distance observed between the level surfaces is considered to be the height difference between the two points.

4.2 Earth Ellipsoid and Normal Ellipsoid

4.2.1 Earth Ellipsoid

The geoid is an irregular surface that approximates the shape of the Earth. Its undulations are mainly generated by the inhomogeneous distribution of mass within the Earth's crust. Given the fact that the mass of the Earth's crust is only one sixty-

fifth of the total mass of the Earth, this irregularity is very small. As a consequence, the geoid closely approximates a regular figure in general. Geodetic observations since the eighteenth century have demonstrated that this regular shape is an oblate ellipsoid of rotation with a bulge at the equator and flattening at the North and South Poles.

A rotational ellipsoid is a geometric figure obtained by rotating an ellipse around its minor axis. Figure 4.8 is an ellipsoid centered at the origin O rotating around its NS axis.

In geodesy, the ellipsoid of rotation that represents the Earth's shape and size is referred to as the Earth ellipsoid, shortened to ellipsoid. The Earth ellipsoid is specified by four parameters: the semimajor axis a and flattening f that represent geometric properties of the Earth; the total mass M of the ellipsoid, which represents the physical properties of the Earth; and the angular velocity ω of the ellipsoid rotating around its minor axis.

Before the 1950s, the geometric parameters of the Earth ellipsoid a and f were computed using data from astronomical, geodetic, and gravimetric observations in particular regions. Results were of low precision and could only represent the geometric figure of some particular regions on the Earth. Since the 1960s, the four geometric and physical parameters of the Earth ellipsoid have been calculated using data from global terrestrial geodetic measurements and satellite geodetic surveys with an increase in precision of two orders of magnitude compared with that prior to the 1950s. One case in point is Ellipsoid GRS80 (Moritz 2000); the difference of a is less than 2 m and the relative mean squared errors of f and GM are $\pm 3 \times 10^{-6}$ and $\pm 2 \times 10^{-7}$, respectively (G is gravitational constant, M is total mass of the Earth). Table 4.1 gives the ellipsoid parameters used in China. The Beijing Geodetic Coordinate System 1954 adopted the Krassowski Ellipsoid, the Xi'an Geodetic Coordinate System 1980 adopted the Ellipsoid GRS 75, and the China Geodetic Coordinate System 2000 (CGCS2000) basically adopted the Ellipsoid GRS80 (GM value of the Ellipsoid GRS80 is defined more precisely). The parameters adopted by the two international geodetic coordinate systems WGS84 (World Geodetic System 1984) and the ITRF (International Terrestrial Reference Frame; see Chap. 7) are the WGS84 Ellipsoid and the GRS 80 Ellipsoid, respectively.

On the Earth ellipsoid, the plane that contains the rotation axis (minor axis) of the reference ellipsoid is called the geodetic meridian plane. The geodetic meridian is the intersection of the plane containing the rotation axis with the surface of the ellipsoid. The plane through the center of the ellipsoid and perpendicular to the axis of rotation is the Earth's equatorial plane. The equator is the intersection of the equatorial plane with the ellipsoid. A parallel circle (parallel line) is an intersection of the plane parallel to the equator with the ellipsoid, also termed circle of latitude. The northernmost point N of the spin axis on the Earth is the North Pole, lying diametrically opposite the South Pole, S.

The Earth's gravity field and its level surface become rather complex given the irregular shape of the Earth and the uneven distribution of mass. To facilitate the study of gravity and the gravity field, the Earth ellipsoid is introduced, which is

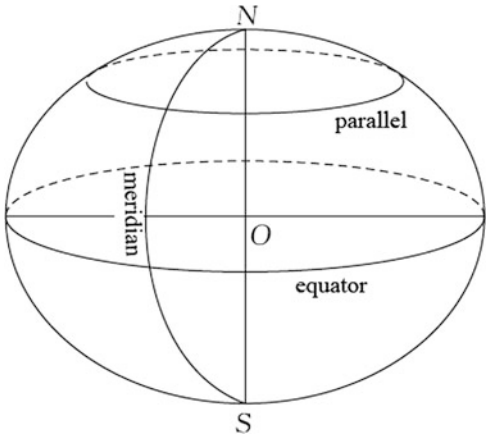


Fig. 4.8 Earth ellipsoid

Table 4.1 Earth ellipsoid parameters (semimajor axis a, flattening f, gravitational constant × total mass GM, and angular velocity ω)

Name of the ellipsoid	Year	a (m)	f	GM ($\times 10^{14}$ m^3/s^2)	ω ($\times 10^{-5}$ rad/s)
Krassowski Ellipsoid	1940	6,378,245	1/298.3	–	–
GRS75	1975	6,378,140	1/298.257	3.986005	7.292115
GRS80	1980	6,378,137	1/298.257222101	3.986005	7.292115
WGS84	1996	6,378,137	1/298.257223563	3.986004418	7.292115
CGCS2000	2008	6,378,137	1/298.257222101	3.986004418	7.292115

called the normal ellipsoid. Due to the irregularity of the actual Earth's shape, a regular surface should be chosen as the reference surface on which geodetic observations and computations are performed. The Earth ellipsoid introduced hereby is called the reference ellipsoid (see Sect. 5.2).

4.2.2 Normal Ellipsoid and Normal Gravity

The normal ellipsoid is an imaginary rotational ellipsoid with regular shape and homogeneous mass distribution that satisfies certain conditions. It is the regular shape of the geoid and is used to represent the ideal body of the Earth. The gravity field generated by the normal ellipsoid is termed the normal gravity field. Corresponding gravity, gravity potential, and level surface are called normal gravity, normal gravity potential, and the spheropotential surface (spherop), respectively. The normal ellipsoid is artificially chosen. The normal gravity potential on the normal ellipsoid is specified as being constant, and its value is equal to the gravity potential W_0 on the geoid (Fig. 4.7).

The normal gravity field is a close approximation to the actual Earth's gravity field. In order to narrow down the difference between the two, we select the normal ellipsoid in accordance with the requirements below:

1. The spin axis of the normal ellipsoid coincides with the Earth's axis of rotation, and with equivalent angular velocity.
2. The center of the normal ellipsoid is at the Earth's center of mass. The coordinate axis coincides with the Earth's principal axis of inertia.
3. The total mass of the normal ellipsoid is equal to that of the actual Earth.
4. The sum of squares of the deviations of the geoid from the normal ellipsoid is the least.

The normal ellipsoid is defined by these four basic parameters: semimajor axis of the ellipsoid a, flattening f, total mass M of the ellipsoid, and the angular velocity ω of the ellipsoid rotating about the minor axis. The former two parameters specify the geometric shape of the ellipsoid and the latter two identify the physical properties of the ellipsoid.

The normal ellipsoid is regular, so obviously its gravitational potential is independent of λ and is only a function of ρ and θ in (4.17). The gravitational potential of the normal ellipsoid is symmetric with respect to the equator. Find the cosine of θ and $180 - \theta$ for the two points that are symmetric to the equator with opposite signs. Therefore, in the spherical harmonics series expansion of the Earth's gravitational potential, there are only even zonal harmonics. Hence, the gravitational potential V of the normal ellipsoid at an exterior point can be obtained from (4.19), given by:

$$V_{(\rho,\theta)} = \frac{GM}{\rho}\left[1 - \sum_{n=1}^{\infty} J_{2n}\left(\frac{a}{\rho}\right)^{2n} P_{2n}(\cos\theta)\right]. \tag{4.23}$$

Equation (4.23) can be determined because J_{2n} is the constant coefficient related to the normal ellipsoid parameters.

Normal gravity can be obtained from the derivative of the normal gravity potential due to their relations. With deviations omitted, the formula for normal gravity value γ_0 on the normal ellipsoid is simplified as:

$$\gamma_0 = \gamma_a\left(1 + \beta \sin^2 B - \beta_1 \sin^2 2B\right), \tag{4.24}$$

where γ_a is the value of gravity at the equator, B is the geodetic latitude of the computation point, and coefficients β, β_1, and equatorial gravity γ_a are given by:

$$\begin{cases} \gamma_a = \dfrac{GM}{ab}\left(1 - \dfrac{3}{2}m - \dfrac{3}{7}mf - \dfrac{125}{294}mf^2\right), \\ \beta = -f + \dfrac{5}{2}m - \dfrac{17}{14}mf + \dfrac{15}{4}m^2, \\ \beta_1 = -\dfrac{1}{8}f^2 + \dfrac{5}{8}mf, \end{cases}$$

where $m = \dfrac{\omega^2 a^2 b}{GM}$, b denotes the semiminor axis of the ellipsoid.

For the normal gravity γ_0 (unit: m/s^2)on the ellipsoid surface of CGCS2000 (cf. Table 4.1), when the permissible error is less than 0.1×10^{-5} m/s^2, (4.24) yields:

$$\gamma_0 = 9.7803253349\left(1 + 0.00530244\sin^2 B - 0.00000582\sin^2 2B\right).$$

The precise formula is:

$$\begin{aligned}\gamma_0 = \gamma_e\big(&1 + 0.005279042982\sin^2 B + 0.000023271800\sin^4 B + 0.000000126218\sin^6 B \\ &+0.000000000730\sin^8 B + 0.000000000004\sin^{10} B\big).\end{aligned}$$

The error in this equation is 0.001×10^{-8} m/s^2.

At the point where the height above the normal ellipsoid is H, the normal gravity value γ is approximately:

$$\gamma = \gamma_0 - 0.3086H. \tag{4.25}$$

This means that when the height increases by 1 m, the normal gravity will decrease by 0.3 mGal (0.3×10^{-5} m/s^2).

The normal gravity γ at any arbitrary exterior point on the CGCS2000 ellipsoid can be computed using the series:

$$\begin{aligned}\gamma = \gamma_0 &- \left(3.08338788871 \times 10^{-6} + 4.429743963 \times 10^{-9}\cos^2 B - 1.9964614 \times 10^{-11}\cos^4 B\right)H \\ &+ \left(7.2442777999 \times 10^{-13} + 2.116062 \times 10^{-15}\cos^2 B - 3.34306 \times 10^{-17}\cos^4 B\right. \\ &\left. - 1.908 \times 10^{-19}\cos^6 B - 4.86 \times 10^{-22}\cos^8 B\right)H^2 \\ &- \left(1.51124922 \times 10^{-19} + 1.148624 \times 10^{-21}\cos^2 B\right. \\ &\left. + 1.4975 \times 10^{-23}\cos^4 B + 1.66 \times 10^{-25}\cos^6 B\right)H^3 \\ &+ \left(2.95239 \times 10^{-26} + 4.167 \times 10^{-28}\cos^2 B\right)H^4,\end{aligned}$$

where γ_0 is measured in meters per second squared, and H is measured in meters. This formula is applied to calculate the error of normal gravity. When H amounts to 20 km, the error is less than 0.1×10^{-8} m/s^2; when H is up to 70 km, the error is less than 1×10^{-8} m/s^2.

There will inevitably be two gravity values that correspond to the gravity field of the actual Earth and that of the normal ellipsoid, namely the true value of gravity g and normal gravity value γ, the difference in value between g and γ, i.e., $g - \gamma$ is the gravity anomaly.

4.2.3 Disturbing Potential

After introducing the normal ellipsoid, there exist two values of gravity potential for any arbitrary points on the Earth: the true (measured) potential W of the real Earth and the normal gravity potential U. There is a difference in value between the two potentials. Subtracting the normal gravity potential from the true potential W of the real Earth, the result is defined as T, the disturbing potential. Namely:

$$T = W - U,$$

or

$$W = U + T, \tag{4.26}$$

which means that the Earth's gravity potential is equal to the sum of the normal gravity potential and the disturbing potential.

It follows from (4.13) that:

$$T = V_E - V_N + Q_E - Q_N.$$

The subscript E is for the actual Earth, and the subscript N for the normal ellipsoid. When the normal ellipsoid is chosen, its axis of rotation is made coincident with the spin axis of the actual Earth and with equal angular velocity; thus $Q_E = Q_N$. It follows that:

$$T = V_E - V_N. \tag{4.27}$$

That is to say, the disturbing potential can be interpreted as the difference in gravitational potential caused by the differences in mass distribution between the Earth and the normal ellipsoid. The mass difference between the Earth and the normal ellipsoid is called the disturbing mass (where the mass refers to the mass difference of each point rather than the difference in total mass). Consequently, the disturbing potential is the gravitational potential of the body constituted by the disturbing masses. The gravitational potential of a body can be expressed as a spherical harmonics series.

From (4.27) we can see that the disturbing potential can be obtained by subtracting the normal gravitational potential from the Earth's gravitational potential. Equation (4.19) describing the Earth's gravitational potential is obtained under

the conditions that the origin of coordinates is at the Earth's mass center; the coordinate axis coincides with the Earth's three principal axes of inertia. Equation (4.23) of the normal gravitational potential, on the other hand, is valid when the coordinate origin is at the center of the ellipsoid; one coordinate axis coincides with the axis of rotation of the normal ellipsoid. When the normal ellipsoid is selected, its center is made at the Earth's mass center, the coordinate axis is made coincident with the Earth's principal axis of inertia, the Earth's axis of rotation coincides with the spin axis of the normal ellipsoid, and the total mass of the normal ellipsoid is equal to that of the Earth. Hence, the disturbing potential T is:

$$T = -\frac{GM}{\rho}\sum_{n=2}^{\infty}\left(\frac{a}{\rho}\right)^n\sum_{k=0}^{n}\left(a'_{nk}\cos k\lambda + b_{nk}\sin k\lambda\right)P_{nk}(\cos\theta), \tag{4.28}$$

where a'_{nk} is the difference in the expansion coefficients between the Earth's gravitational potential and the normal gravitational potential.

4.3 Height Systems

4.3.1 Requirements for Selecting Height Systems

The height of a point on the Earth's surface can be determined by leveling, trigonometric leveling, and GPS measurement. Whichever method is used, a reference surface (zero-elevation surface) and reference line (the line along which the height is measured) will be involved. The height of a point on the Earth's surface is geometrically defined as the distance from the point along the reference line to the reference surface. Different reference lines or reference surfaces for heights will constitute different height systems. Obviously, the height of the same Earth's surface point in different height systems also varies.

For the height system to be chosen, the following requirements need to be fulfilled:

1. To represent the position of a point, the height of the point is required to be unambiguous and independent of the leveling path.
2. In practice, when converted to the adopted height system, the corrections to the measured height differences for points in a limited area should be very small so that they can possibly be ignored while dealing with low-order leveling data.
3. From the geometric problem-solving perspective, the ellipsoidal height is the sum of the measured height and the geoid height; thus it requires that the adopted height system should make the method for determining the difference between the geoid and the reference ellipsoid (normal ellipsoid) sufficiently rigorous and convenient, as well as practical.

4. From the physical problem-solving perspective, the chosen height system is also required to ensure that the height of each point on the same level surface be equivalent as much as possible. This is because the leveling data is actually used to determine the physical problem of the relative position of the Earth's natural surface and the level surface of the real gravity field, which is essential in avoiding the "water runs uphill" phenomenon in engineering application.

Requirements 2 and 4 are in fact contradictory. Therefore, in practice, choosing the most appropriate height system requires compromise according to the different requirements of the application.

4.3.2 Non-uniqueness of Leveled Height

The principle of leveling is based on the fact that level surfaces are parallel to each other. In relatively small areas, the level surfaces on which the fore and back rod are placed are viewed as parallel to each other. The measured distance between the two level surfaces is the height difference between the two points. It is known that level surfaces are not generally parallel. When the leveling lines are of greater lengths in larger measured areas, the level surfaces cannot be considered as parallel, meaning that the influence of non-parallel level surfaces on leveled heights must be considered.

As shown in Fig. 4.9, suppose that the connecting surface between points O, E, and C is the geoid (reference surface for heights); then the observed height of point B can be derived by summing the measured height differences Δh_1, Δh_2, . . .along the leveling line OAB at each station:

$$H_M^B = \Delta h_1 + \Delta h_2 + \cdots = \sum_{OAB} \Delta h.$$

The subscript M indicates measured height.

The height of point B can also be calculated by summing the leveled height differences $\Delta h_1^{'}$, $\Delta h_2^{'}$, $\cdots$ along leveling line ONB:

$$H_M^{'B} = \Delta h_1^{'} + \Delta h_2^{'} + \cdots = \sum_{ONB} \Delta h^{'}$$

Due to the unparallel nature of level surfaces, the corresponding height differences Δh_i and $\Delta h_i^{'}$ are not equal; thus H_M^B and $H_M^{'B}$ are not equal either.

In Fig. 4.9, $OABNO$ is a closed leveling loop, and obviously:

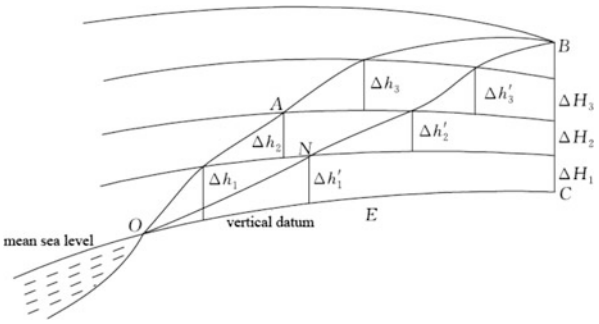


Fig. 4.9 Effects of non-parallel level surface on leveling

$$\sum_{OAB} \Delta h \neq \sum_{BNO} \Delta h',$$
$$\sum_{OAB} \Delta h + \sum_{BNO} \Delta h' = w \neq 0. \tag{4.29}$$

So, similarly to errorless leveling, the error of closure of leveling loops w still cannot be zero. Such closing error of leveling loops caused by unparallel level surfaces is called theoretical misclosure.

The height of a point should be single-valued and be independent of the leveling routes. So, in dealing with leveling results, the properties of the theory of the Earth's shape must be taken into consideration; the height system must be defined reasonably and corrections applied. These corrections should be very small so that they could be neglected in dealing with low-order leveling results.

4.3.3 Orthometric Height

Referenced to the geoid, the orthometric height (H_o) is the length between the geoid (reference surface) and a point on the Earth's surface measured along the plumb line. As shown in Fig. 4.9, the height differences between each level surface measured from point B along the plumb line is expressed by ΔH, then H_O^B, the orthometric height of point B is:

$$H_O^B = \Delta H_1 + \Delta H_2 + \cdots = \sum_{CB} \Delta H = \int_{CB} \mathrm{d}H. \tag{4.30}$$

Since the level surface is an equipotential surface, the potential energy difference between two infinite close level surfaces in Fig. 4.9 is given by:

$$\begin{aligned} g\mathrm{d}h &= g^B \mathrm{d}H \\ \mathrm{d}H &= \frac{g}{g^B}\mathrm{d}h, \end{aligned} \tag{4.31}$$

where g is the gravity at dh along the leveling line and g^B is the gravity at dH along the plumb line from point B. Substituting (4.31) into (4.30), we have:

$$H_O^B = \int_{CB} \mathrm{d}H = \int_{OAB} \frac{g}{g^B}\mathrm{d}h. \tag{4.32}$$

The gravity along the plumb line, g^B, varies with depth. Let their average be g_m^B; then:

$$H_O^B = \frac{1}{g_m^B}\int_{OAB} g\mathrm{d}h, \tag{4.33}$$

where g_m^B is a certain value relative to a certain surface point. $\int g\mathrm{d}h$, independent of the leveling path, is the potential energy difference between the level surface passing through point B and the geoid. So, the orthometric height is a unique value. However, since g_m^B is the average gravity at depth and we cannot know for sure the subsurface mass density distribution, g_m^B therefore can neither be measured nor precisely calculated. Thus, the orthometric height of a point can only be approximately evaluated.

4.3.4 Normal Height

The reason why orthometric height cannot be precisely obtained lies in that g_m^B of point B cannot be measured accurately. Replacing g_m^B in (4.33) with the normal gravity γ_m^B, one can get the normal height that belongs to another height system, denoted by H_N, namely:

$$H_N^B = \frac{1}{\gamma_m^B}\int_{OAB} g\mathrm{d}h, \tag{4.34}$$

where g can be measured through gravimetry along the leveling line, dh can be measured by leveling, and γ_m^B can be calculated by the normal gravity formulae (4.24) and (4.25). Thus, we can get a precise normal height whose value is unique, without varying with the changes in leveling routes. The concept of normal heights was formulated by the Russian geodesist M.S. Molodensky in 1945. In China, the normal height system is adopted as the unified system for computing the height of a point on the Earth's surface.

If the normal height of each surface point is H_N, measuring H_N downward along the plumb line (the normal gravity line in fact) results in the corresponding points of each surface point. A continuous curved surface as the reference surface for normal heights can be formed by connecting these corresponding points. It is also called the quasi-geoid because of its close approximation to the geoid. Therefore, the so-called normal height system is the height system with the quasi-geoid as its reference surface. The normal height of a surface point is the distance from this point to the quasi-geoid along the plumb line (the normal gravity line).

As an auxiliary surface for calculation, the quasi-geoid approximates, but does not equal, the geoid. It has no strict geometric or physical meanings.

The difference between the quasi-geoid and the geoid (i.e., the difference between the orthometric height and the normal height) is associated with the height of a point and the mass distribution inside the Earth. Neglecting the sea surface topography, at mean sea level the observed height difference $\mathrm{d}h = 0$, so $H_N = H_O = 0$; that is, the quasi-geoid coincides with the geoid on the oceans. So, the height origin as the reference surface for heights is applicable to both the quasi-geoid and the geoid. In plain areas, the difference between the quasi-geoid and the geoid is a few centimeters whereas in mountainous regions it can reach values of about 3 m.

In real applications, using (4.34) to calculate the normal height is not convenient. Considering that the actually measured gravity value is made up of two components, the normal gravity γ and the gravity anomaly $(g - \gamma)$, the corresponding normal height can be calculated by adding the observed height difference for each segment of leveling and the correction to non-parallel spheropotential surfaces and the gravity anomaly correction. Omitting derivations, the result is:

$$H_N^B = \int_{OAB} \mathrm{d}h + \frac{1}{\gamma_m^B} \int_{OAB} \left(\gamma_0 - \gamma_0^B\right)\mathrm{d}h + \frac{1}{\gamma_m^B} \int_{OAB} (g - \gamma)\mathrm{d}h. \tag{4.35}$$

where the meanings of each term on the right side of the equation are as follows:

The first term is the leveled height difference.

In the second term, γ_0 is the normal gravity of each point along the leveling line OAB. Since the spheropotential surfaces are also not parallel and vary with latitude, $\gamma_0 \neq \gamma_0^B$, this term is called the correction to the non-parallel spheropotential surface.

In the third term, $(g - \gamma)$ is the gravity anomaly, resulting from the inconsistency between the spheropotential surface (spherops) and the geopotential surface (geops).

4.3.5 *Dynamic Height*

A level surface is equipotential, on which the gravity potential of each point is equal, but the orthometric height and normal height of each point can be different. Assume that point A and point B are on the same level surface, then:

$$\int_{OA} g\mathrm{d}h = \int_{OB} g\mathrm{d}h. \tag{4.36}$$

because

$$g_m^A \neq g_m^B,$$
$$\gamma_m^A \neq \gamma_m^B.$$

Thus it can be seen from (4.33) and (4.34) that:

$$H_O^A \neq H_O^B,$$
$$H_N^A \neq H_N^B.$$

This would cause inconvenience to the design and construction of large-scale water conservancy projects. A new height system is therefore needed to make "the height of each point on the same level surface equal." Hence, a dynamic height system is usually adopted in a water conservancy project. The dynamic height (H_D) of point B is:

$$H_D^B = \frac{1}{\gamma_{45}^B}\int g\mathrm{d}h. \tag{4.37}$$

The above equation shows that the dynamic height is obtained by replacing γ_m in (4.34) with γ_{45}, the normal gravity at latitude 45°. Points on the same level surface have the same dynamic height. So the dynamic height system is a system where the dynamic height of each point is represented by the normal height (H_N) of the point at 45° latitude.

"Local dynamic height system" is adopted by some authorities to bring the dynamic height over an area closer to the normal height in the same area, that is:

$$H_{LD} = \frac{1}{\gamma_{\varphi m}}\int g\mathrm{d}h, \tag{4.38}$$

where $\gamma_{\varphi m}$ is the normal gravity value at mean latitude φ_m of the survey area.

The dynamic height and the normal height of a surface point can be easily interconverted. From (4.34) and (4.36), we have:

$$\gamma_m^B H_N^B = \gamma_{45}^B H_D^B,$$
$$H_N^B = H_D^B - \frac{\gamma_m^B - \gamma_{45}^B}{\gamma_m^B} H_D^B. \tag{4.39}$$

4.3.6 Geopotential Number

The height of a point on the Earth's surface is represented by the difference between the potential of the geoid W_0 and that of the equipotential surface passing through this point W, also known as the geopotential number, namely:

$$C = W_0 - W = \int_{OAB} g\mathrm{d}h, \tag{4.40}$$

where OAB is the level line (see Fig. 4.9), dh is the difference in height measured during each setup of leveling, and g is the mean value of gravity along the leveling lines. With the geoid as the reference surface, geopotential numbers are not measured in meters but in potential differences kGal m (10^5 cm^2/s^2). On the same level surface, the geopotential number of every point is equal and its value can be obtained by multiplying dh by the mean value of gravity (g) for the setup. The leveling results expressed by geopotential numbers can be conveniently converted to the orthometric height, normal height, and the dynamic height.

The orthometric, normal, and dynamic height systems all have their respective advantages and disadvantages. The coexistence of them not only makes the height systems non-unified, but also increases the difficulty in combined processing of the leveling data. Obviously, all three height systems have one part in common, the geopotential number $\int g\mathrm{d}h$, which is the potential energy that the elevation point possesses from the location of the point to the geoid. Its relationship with the three heights is simple and clear; thus, the leveling data can be processed using geopotential numbers with a unified height.

Although the geopotential number does not have the dimension of a length, it can be considered as a natural measure for height.

4.3.7 Geodetic Height

Geodetic height (ellipsoidal height) refers to the reference ellipsoid, measured along the ellipsoidal normal (see Sect. 5.2). In modern geodesy, the reference ellipsoid surface can be seen as the normal ellipsoid surface because the reference ellipsoid is consistent with the normal ellipsoid. The distance from a surface point to the reference ellipsoid along the ellipsoidal normal is defined as the geodetic height of this point.

As shown in Fig. 4.10, point P is a point on the Earth's surface. Its projection onto the ellipsoid along the ellipsoidal normal is P_0, and the distance $\overline{PP_0}$ is the geodetic height H.

The geodetic height or geodetic height difference between two points on the Earth's surface can be obtained by satellite positioning survey or trigonometric leveling. Given the geodetic height of one point, that of the other can be easily

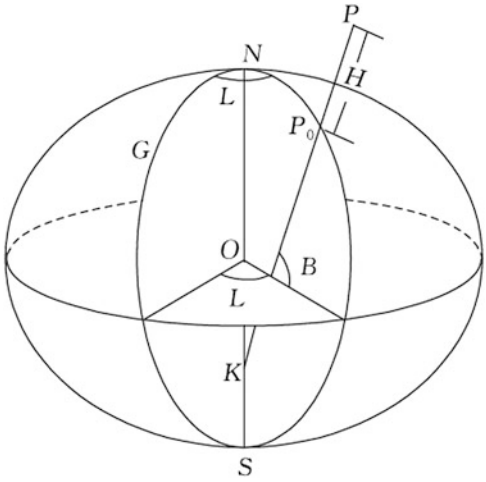


Fig. 4.10 Geodetic height

calculated. The leveled height, with corrections applied, can be converted to the geodetic height.

4.4 Relationship and Transformation Between Different Height Systems

4.4.1 Relationship Between Orthometric Height, Normal Height, and Geodetic Height

From the previous discussion, we know that the same surface point has at least five different values of height due to different reference surfaces for heights. In other words, height is relative to a reference surface. Its precision depends on the precision of both the observed quantity and the reference surface used. The relationship between different reference surfaces for heights will now be described.

The Earth's surface points project on the ellipsoid in two ways: the Helmert's projection and Pizzett's projection, as shown in Fig. 4.11.

According to Helmert's projection, a point P on the Earth's surface is directly projected onto the ellipsoid along the normal ($P_0P = H$). In Pizzett's projection, the same point P is projected along the plumb line onto the geoid ($P'P = H_O$). The point on the geoid is then projected once again along the normal to the ellipsoid ($P_0{}'P' = N$).

As is known, the geoid and the reference ellipsoid surface are usually neither coincident with nor parallel to each other, so there is an angle μ between the plumb line and the normal, called the deflection of the vertical. Using these two different projection methods will achieve different results, although the difference is quite

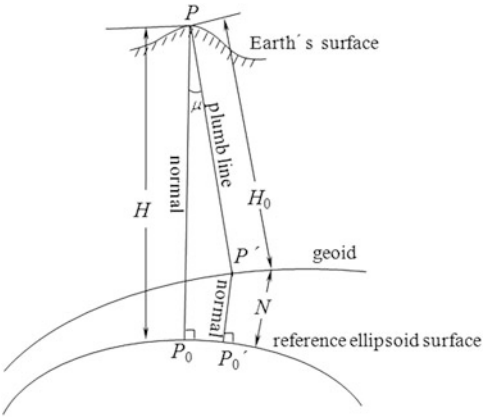


Fig. 4.11 Helmert's projection (P–P_0) and Pitzzetti's projection (P–P'–P_0')

slight. Suppose μ=60″, H = 1,000 m (in most cases, the values are much smaller than the values given here), then the difference between H and $H_O + N$ is only 0.1 mm. The distance of $P_0P_0^{'}$ is merely 30 cm, whose effect on the geodetic longitude and latitude is only 0.01″, far less than the 0.3″ error of the astronomical measurement (the astronomical longitude and latitude corresponds to points on the geoid). Therefore, difference between these two projection methods can be ignored in practical cases.

The geodetic height of a point on the Earth's surface can be obtained directly using GPS measurement, and the relationship between the surface point and its corresponding point projected onto the ellipsoid is established by the Helmert's projection. However, in classical geodetic survey, the ellipsoidal height is not directly measured, but calculated using orthometric height (or normal height) with corrections applied. Therefore, using Pizzett's projection to determine the corresponding relationship between a surface point and its projection point on the ellipsoid surface is theoretically rigorous. However, in practice, since the difference between Helmert's and Pizzett's projections can be ignored and Helmert's projection avoids double projection, first to the geoid and then to the ellipsoid, which is more convenient in application, Helmert's projection is adopted in classical geodetic computations.

So, as shown in Fig. 4.11, the geodetic height can be calculated according to:

$$H = H_{\mathrm{O}} + N, \tag{4.41}$$

where H_{O} is the orthometric height and N is the distance from the geoid to the reference ellipsoid, called the geoid height (geoid undulation or geoid separation).

China adopts the normal height system and the geodetic height can be calculated according to:

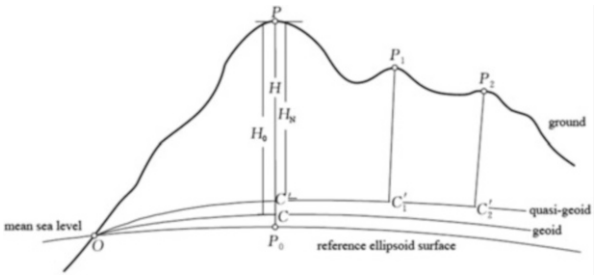


Fig. 4.12 Reference ellipsoid, geoid, and quasi-geoid

$$H = H_N + \zeta, \tag{4.42}$$

where H_N is the normal height and ζ is the distance from the quasi-geoid to the reference ellipsoid, called the height anomaly.

Figure 4.12 illustrates the relationship between the reference ellipsoid, the geoid, and the quasi-geoid and their corresponding geodetic height, orthometric height, and normal height.

4.4.2 Determination of Height Anomaly or Geoid Height

As pointed out previously, the geodetic height of a surface point consists of the normal height and the height anomaly. Given the geodetic height and the normal height of a point, the height anomaly can be computed from the difference between the two, namely:

$$\zeta = H - H_N. \tag{4.43}$$

Using GPS measurements, the geodetic longitude L and latitude B and the geodetic height H of a surface point can be determined precisely. If leveling is also carried out on the GPS point (this point is called the GPS-leveling point), then the normal height H_N of this point can be calculated and the height anomaly of this point can be determined by (4.43).

By setting a few GPS-leveling points in a certain region, several discrete ζ values of this region can be determined, and thus the quasi-geoid of this region can be fitted through a mathematical method (i.e., deducing the height anomaly of an unknown point). Such a method for deducing the height anomaly is called the GPS leveling method. A variety of mathematical methods are used in GPS leveling, such as the polynomial fitting method, polyhedral function fitting, the moving surface method, the weighted average method, the collocation method, etc. In real applications, GPS leveling and gravity data are usually used for a combined solution. Here, we will only present the basics of GPS leveling rather than provide a thorough review of the

good and bad points of each method. Polynomial fitting will be used to explain the basics of GPS leveling because it is readily understandable.

If the quadratic polynomial is used as the fitting model of the height anomaly, then the height anomaly of this region ζ can be expressed by:

$$\zeta = \alpha_0 + \alpha_1 \Delta L + \alpha_2 \Delta B + \alpha_3 \Delta L^2 + \alpha_4 \Delta L \Delta B + \alpha_5 \Delta B^2, \tag{4.44}$$

where $\Delta L = L - L_0$ and $\Delta B = B - B_0$ are the differences between the geodetic longitude and latitude L, B of the point to be calculated and L_0, B_0 of a known point in this region; $\alpha_i (i = 0 \cdots 5)$ is the undetermined coefficient. Suppose the number of GPS-leveling points in this region is n, meaning that $\varsigma_i (i = 1 \cdots n, n \geq 6)$ is already known; then n equations can be formed by (4.44). Let:

$$\vec{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} \quad \mathrm{X} = \begin{pmatrix} 1 & \Delta L_1 & \Delta B_1 & \Delta L_1^2 & \Delta L_1 \Delta B_1 & \Delta B_1^2 \\ 1 & \Delta L_2 & \Delta B_2 & \Delta L_2^2 & \Delta L_2 \Delta B_2 & \Delta B_2^2 \\ 1 & \Delta L_3 & \Delta B_3 & \Delta L_3^2 & \Delta L_3 \Delta B_3 & \Delta B_3^2 \\ 1 & \Delta L_4 & \Delta B_4 & \Delta L_4^2 & \Delta L_4 \Delta B_4 & \Delta B_4^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta L_n & \Delta B_n & \Delta L_n^2 & \Delta L_n \Delta B_n & \Delta B_n^2 \end{pmatrix} \quad \vec{\zeta} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \vdots \\ \zeta_n \end{pmatrix}. \tag{4.45}$$

Then the least square solution of the undetermined coefficient vector $\vec{\alpha}$ is:

$$\vec{\alpha} = \left(\mathbf{X}^{\mathrm{T}} \mathbf{P} \mathbf{X}\right)^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{P} \vec{\zeta}, \tag{4.46}$$

where $\mathbf{P}$ is the weight matrix. If the known data are considered as mutually independent, then the principal diagonal element of $\mathbf{P}$ is the weight of $\varsigma_i (i = 1 \cdots n)$.

After the determination of the undetermined coefficient, the value of ζ of the unknown point can be calculated according to L, B of this point by (4.44).

This method can also be used to determine the geoid height N by replacing H_{N} with the orthometric height H_{O} of the GPS-leveling point and ζ with N in (4.43).

At present, centimeter-level or even higher-level precision can be achieved in determining the height anomaly by the method of GPS leveling combined with gravity data.

From the above, it is known that GPS leveling can determine the geoid directly on land. On the ocean, the geoid can be deduced directly by satellite altimetry (see Chap. 2). The shape of the geoid can be deduced directly now, whereas it could only be obtained by gravity field in the period of classical geodetic survey.

The separation of geoid from ellipsoid N (geoid height) and the height anomaly ζ can also be determined based on the Earth gravity field model. Put more directly, the location of the geoid is related to the gravity potential, whereas the location of

the ellipsoid is related to the normal gravity potential; thus N or ζ can be determined by the disturbing potential T. With derivations omitted, the formula is given by:

$$N = \frac{T_0}{\gamma_0}, \tag{4.47}$$

where T_0 is the disturbing potential on the geoid and γ_0 is the normal gravity value on the normal ellipsoid. Similarly, we have:

$$\zeta = \frac{T_p}{\gamma_m}, \tag{4.48}$$

where T_p is the disturbing potential at the Earth's surface and γ_m is the integral mean of the normal gravity along line segment P_0P (see Fig. 4.12).

4.4.3 Grid Models of Height Anomaly or Geoid Height

The grid model of height anomaly is a discrete numerical representation of height anomaly within a certain range. It is an aggregation of height anomaly values of all the equidistant grid points within this range, stored in the database in the form of graticule data structure. This structure divides the geographical area covered by the database into regular trapezoid grids based on longitude and latitude lines (see Fig. 4.13) and uses the grid area or the vertical and horizontal intersection of grid lines as the node to store the corresponding height anomaly values at the intersection or the average height anomaly values within the grid area.

For example, dividing a 1:1,000,000 map sheet into 240 × 360 equidistant trapezoid grids based on directions in which the longitude and latitude lines run according to a $1' \times 1'$ field range, each grid can be numbered as 00001–86400, from left to right by rows and from top to bottom by columns (similar to the international subdividing method).

A multilevel grid structure can be used in grid models. Grids at different levels have different distance intervals. The smaller the grid, the more precise the continuous height anomaly can be. For example, we can divide the database building area into different regions according to the known data distribution, and different regions use grids with different distance intervals.

To establish the grid model of height anomaly over an area is to divide this area into equidistant regular grid cells according to the longitude and latitude lines, and then calculate the height anomaly of each grid node using a certain method according to the known height anomaly within the area (e.g., GPS-leveling points).

When using a height anomaly grid model, one first needs to identify the grid cell in which the point lies and then use bilinear interpolation according to the height anomaly of the four grid intersections to calculate the height anomaly of the point.

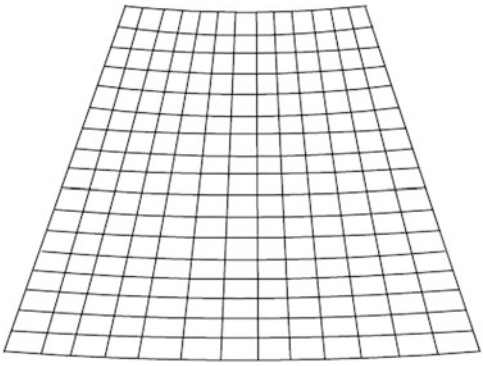


Fig. 4.13 Grids based on longitude and latitude lines

Normal height is needed in topographic mapping and engineering construction. Traditional leveling is labor-consuming and inefficient, whereas GPS leveling can provide the ellipsoidal height. The grid model of height anomaly should be known while using GPS measurements to replace leveling, and then the ellipsoidal height can be converted into the normal height. At present, leveling has gradually been replaced by GPS measurements in height connection surveys for aerial photography, project completion surveys, pipeline surveys, etc. Therefore, the establishment and refinement of the grid model of height anomalies have obvious economic benefits.

The grid model of geoid height is established in the same way as that of the height anomaly.

Review and Study Questions

1. Derive the formula for the gravitation of a body of mass.
2. What is the potential function of force?
3. Prove that the gravitational potential of a mass point is $V_{(x,y,z)} = \dfrac{fm}{r}$ and write down the integral formula for the gravitational potential of the Earth.
4. Prove that the centrifugal potential is $Q = \frac{\omega^2}{2}(x^2 + y^2)$.
5. What is harmonic function? Write down the general form of the gravitational potential of the Earth.
6. Explain the physical meaning of zero and first-degree terms of the gravitational potential coefficient.
7. What roles do level surface and plumb line play in field survey?
8. What is the geoid and its role in geodetic survey?
9. Explain the concepts of normal ellipsoid, normal gravity, spheropotential surface, and gravity anomaly.
10. Explain the reason for the existence of the theoretical misclosure of leveling loops.
11. Explain with formulae the concepts of orthometric heights and normal heights.

Chapter 5
Reference Ellipsoid and the Geodetic Coordinate System

The shape of the Earth approximates a regular ellipsoid, which can therefore be used to represent the mathematical shape of the geoid. One can also establish a one-to-one correspondence between the points on the Earth's surface and the points on the ellipsoid that is used as the reference surface. Based on some relevant mathematical properties of the ellipsoid, this chapter discusses the methods for reducing the elements of terrestrial triangulation and trilateration to a reference ellipsoid and establishes the models to transform mutually between the geodetic coordinate system, geodesic polar coordinate system, and geodetic Cartesian coordinate system (geodetic spatial rectangular coordinate system).

5.1 Fundamentals of Spherical Trigonometry

5.1.1 Spherical Triangle

A spherical triangle is a closed figure formed on the surface of a sphere that is bounded by three arcs of great circles. The great circle is defined to be the intersection of a sphere with a plane containing the center of the sphere (Fig. 5.1). The three arcs of great circles are called the sides of the spherical triangle, denoted by lowercase letters a, b, and c. The spherical angles formed by the arcs of great circles are called the angles of the spherical triangle, denoted by the uppercase letters A, B, and C.

A trihedron O-ABC is formed by connecting the vertices of the spherical triangle ABC with the center of the sphere O (Fig. 5.1). The radian measure of a central angle of a circle is equivalent to the length of the arc the angle subtends, which yields:

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_5,

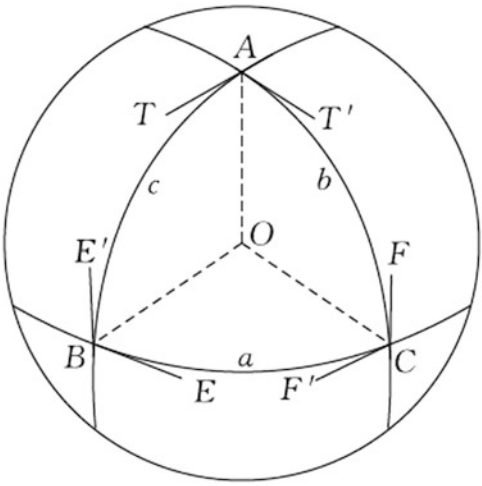


Fig. 5.1 Spherical triangle

$$a = \angle BOC, b = \angle AOC, c = \angle AOB.$$

Given are:

$$A = \angle TAT^{'}, B = \angle EBE^{'}, C = \angle FCF^{'}.$$

Hence, the arc length of any side of the spherical triangle equals its subtended face angle of the trihedral angle; the angles of the spherical triangle are equal to the corresponding dihedral angles of the trihedral angle.

5.1.2 Spherical Excess

Spherical excess is the amount by which the sum of the angles of a spherical triangle exceeds the sum of the angles of a plane triangle, denoted by ε, namely:

$$\varepsilon = A + B + C - 180^{\circ}. \tag{5.1}$$

The computational formula of ε is given by:

$$\varepsilon = \frac{S}{R^2}, \tag{5.2}$$

where S denotes the area of the spherical triangle and R is the radius of the sphere.

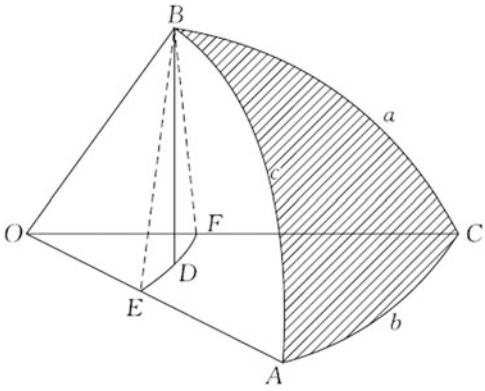


Fig. 5.2 Derivations of the sine formula

5.1.3 *Formulae for Spherical Trigonometry*

The formulae for spherical trigonometry are defined as the formulae applied to obtain the unknown parts based on the given elements (sides, angles) of a spherical triangle.

Sine Formula

In the spherical triangle *ABC* illustrated in Fig. 5.2, the length of the sphere radius is unity and we have:

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}, \tag{5.3}$$

which means that in any spherical triangle, the sines of the sides are proportional to the sines of their opposite angles), proved as follows.

Figure 5.2 is the trihedron *O*-*ABC*, a line *BD* is drawn perpendicular to the plane *OAC* through point *B*. Then, through point *D* draw *DE* and *DF* perpendicular to *OA* and *OC*. Join BE and *BF*; one obtains four right-angled plane triangles *OBE*, *OBF*, *BDE*, and *BDF*. Meanwhile:

$$\angle BOC = a, \angle AOC = b, \angle AOB = c, \angle BED = A, \angle BFD = C.$$

It is given that:

$$\frac{\sin c}{\sin C} = \frac{\frac{BE}{OB}}{\frac{BD}{BF}} = \frac{BE \cdot BF}{OB \cdot BD}, \quad \text{and}$$

$$\frac{\sin a}{\sin A} = \frac{\frac{BF}{OB}}{\frac{BD}{BE}} = \frac{BE \cdot BF}{OB \cdot BD}.$$

Hence:

$$\frac{\sin a}{\sin A} = \frac{\sin c}{\sin C}.$$

Similar results hold for the other sides and angles:

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B}.$$

Combining these two series gives the sine formula.

Other Formulae

Other commonly used formulae are provided below without derivations (see, e.g., Casey 2005).

The law of cosines for sides is:

$$\cos a = \cos b \cos c + \sin b \sin c \cos A \tag{5.4}$$

The law of cosines for angles is:

$$\cos A = -\cos B \cos C + \sin B \sin C \cos a \tag{5.5}$$

The laws of sines and cosines is:

$$\left.\begin{array}{l} \sin a \cos B = \sin c \cos b - \cos c \sin b \cos A \\ \sin a \cos C = \sin b \cos c - \cos b \sin c \cos A \\ \sin A \cos b = \sin C \cos B + \cos C \sin B \cos a \\ \sin A \cos c = \sin B \cos C + \cos B \sin C \cos a \end{array}\right\} \tag{5.6}$$

The laws of cotangents is:

$$\left.\begin{array}{l} \cot a \sin c = \cos c \cos B + \sin B \cot A \\ \cot a \sin b = \cos b \cos C + \sin C \cot A \\ \cot A \sin C = -\cos C \cos b + \sin b \cot a \\ \cot A \sin B = -\cos B \cos c + \sin c \cot a \end{array}\right\} \tag{5.7}$$

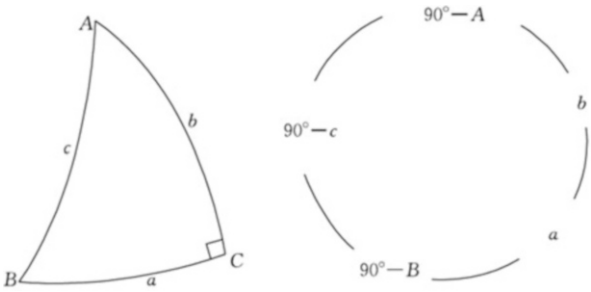


Fig. 5.3 Napier's rules for right-angled spherical triangles

The law of tangents is:

$$\frac{\tan\frac{1}{2}(A+B)}{\tan\frac{1}{2}(A-B)}=\frac{\tan\frac{1}{2}(a+b)}{\tan\frac{1}{2}(a-b)}. \tag{5.8}$$

Napier's Rules for Right-Angled Spherical Triangles

Let one angle of a spherical triangle ABC be 90°, then the cosine of this right angle is 0 and the sine is 1. Substituting into the above formulae, one can obtain the relations between sides and angles of a right-angled spherical triangle. To facilitate memorization, Napier presented some rules.

Except for the right angle C, there are five elements of a spherical triangle ABC arranged in the form of a circle. Keep the two elements adjacent to the right angle C and replace all elements non-adjacent to the right angle C by their complement to 90° (hypotenuse c and angles A and B; see Fig. 5.3). Then Napier's rules hold that the sine of any element in the circle is equal to:

1. The product of the tangents of the adjacent two elements
2. The product of the cosines of the opposite two elements

Take the angle $90^\circ - c$ for example; its adjacent elements are $90^\circ - A$ and $90^\circ - B$, and its opposite elements are a and b. Hence:

$$\sin(90^\circ - c) = \tan(90^\circ - A)\tan(90^\circ - B) \quad \text{and}$$
$$\sin(90^\circ - c) = \cos a \cos b.$$

Namely:

$$\cos c = \cot A \cdot \cot B, \text{and}$$
$$\cos c = \cos a \cdot \cos b.$$

5.2 Reference Ellipsoid

5.2.1 *Reference Surface for Geodetic Surveying Computations*

Section 4.2 introduced the concept of the Earth ellipsoid. It can also be used as the reference surface for geodetic surveying computations, which is called the reference ellipsoid.

Because of the irregularity of the actual shape of the Earth, a regular curved surface should be selected as the reference surface for the performance of geodetic computations. Conventional terrestrial surveys can only determine directions, distances, and astronomical azimuths between points on the Earth's surface, whereas to obtain coordinates of the horizontal control points, a series of computations need to be carried out and a reference surface upon which computations are performed is therefore needed.

The reference surface that fits for the geodetic surveying computations should satisfy the following three conditions:

1. The reference surface should be a curved surface that approximates the physical shape of the Earth, so that the corrections for reduction of the terrestrial observations are small.
2. The curved surface should be a mathematical surface on which computations are easily performed so as to assure the possibility of calculating coordinates through observational quantities.
3. The positions of the curved surface relative to the geoid should be fixed so as to establish the one-to-one correspondence between the points on the Earth's surface and those on the reference surface.

We know that an oblate ellipsoid of rotation approximates the geoid with a bulge at the equator and flattening at the North and South Poles. In fact, precise observations have shown that the North Pole bulges out by 16 m and the South Pole is depressed by approximately 16 m when the geoid is compared with a properly defined ellipsoid (cf. Fig. 5.4). The Earth is thus claimed to be "pear-shaped," which is somewhat exaggerated. This slight difference, however, compared to the difference of 21.4 km between the Earth's equatorial radius and the polar radius is insignificant.

The intersection line between the geoid and the equatorial plane is not a perfect circle, but more closely approximates an ellipsoid. The major axis of the ellipsoid on the equator is at 15° west longitude. The difference between the semimajor axis (equatorial radius) and the semiminor axis (polar radius) is 69.5 m. The equatorial flattening is 1:91,827, which is approximately one three-hundredth of the polar flattening (cf. Fig. 5.5).

As a result, the "pear-shaped" sphere or the triaxial ellipsoid is a mathematical surface that is an approximation to the true shape of the Earth. However, it will be

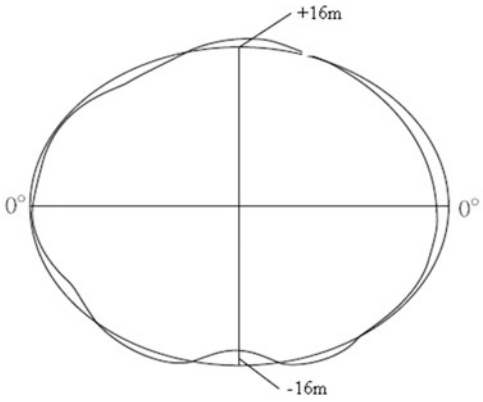


Fig. 5.4 Intersection line between the geoid and the meridian plane ($L = 90°$)

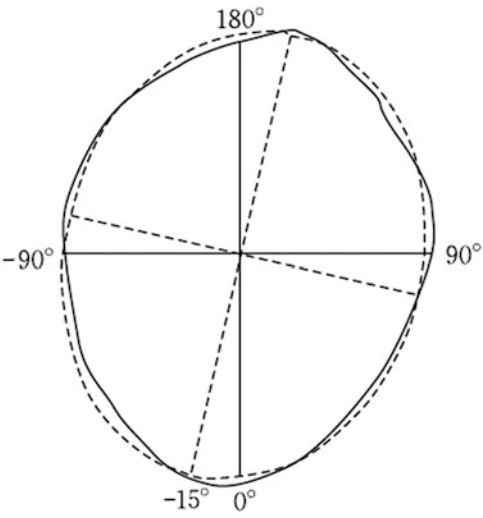


Fig. 5.5 Intersection line between the geoid and the equatorial plane

complicated and inconvenient to perform geodetic surveying computations on this pear-shaped surface. On the other hand, the rotational ellipsoid on which computations are performed provides a tradeoff between computational precision and simplicity and convenience in calculations. Therefore, the ellipsoid of rotation is always chosen as the preferred reference surface for geodetic computations.

When a set of ellipsoidal parameters or an Earth ellipsoid is selected, its location relative to the geoid should be determined, namely to complete the orientation of the ellipsoid. Therefore, the corresponding relationship between the Earth's surface and the ellipsoid can be established to reduce the observations from the terrestrial geodetic control network to the ellipsoid.

The reference ellipsoid is the Earth ellipsoid with defined parameters and orientation. The terrestrial observations in the geodetic control network need to be reduced to the reference ellipsoid and computations are to be performed on this

surface. Hence, the reference ellipsoid becomes the reference surface for surveying computations. The points from the physical surface of the Earth are projected directly onto the ellipsoid along the ellipsoidal normal. As a result, the ellipsoidal normal becomes the datum line for surveying computations.

The reference ellipsoid has defined the geodetic coordinate system (cf. Sect. 5.3). If two countries or one country at different periods of time have adopted different reference ellipsoids, i.e., different geodetic coordinate systems, then one coordinate system needs to be transformed to another in order to make use of each other's results.

The reference ellipsoid is a mathematically defined surface that approximates the physical shape of the Earth. As the reference surface for surveying computations, the reference ellipsoid has played prominent roles in surveying and mapping, as follows:

1. The reference ellipsoid is used as the reference surface for the determination of the horizontal coordinates (geodetic longitude and latitude) and the geodetic height of a point on the Earth's surface.
2. The reference ellipsoid is the reference surface to describe the shape of the geoid. The vertical distance from the geoid to the reference ellipsoid is called geoid undulation. The angle between the plumb line and the surface normal to the reference ellipsoid at a given point is the deflection of the vertical. The geoid undulation and deflection of the vertical reflect the distance between the two surfaces and the deviation of the geoid from the ellipsoid, and also characterize the shape of the geoid.
3. The reference ellipsoid also serves as the reference surface for map projection, where the reference ellipsoid is used to represent the Earth when discussing the correlations between two mathematically defined surfaces.

To study global geodetic problems, there needs to be a reference ellipsoid that best fits the geoid throughout the entire Earth, that is, a general Earth ellipsoid. Its center must coincide with the center of the Earth. If the study is conducted both geometrically and physically, then the general Earth ellipsoid can be defined as the normal ellipsoid that best represents the shape of the geoid.

The normal ellipsoid is the reference surface for studying the Earth's gravity field in physical geodesy. The reference ellipsoid, on the other hand, is the reference surface for studying geodetic computations in geometric geodesy. Practically, due to the same mathematical properties between the normal ellipsoid and the reference ellipsoid, the normal ellipsoid can be used as both the physical and mathematical reference surface in geodesy.

Modern geodesy has realized the geocentric orientation of the reference ellipsoid and has made the reference ellipsoid correspond to the normal ellipsoid. The orientation of the reference ellipsoid established by the classical geodetic techniques can only approximate the geoid of its home country or region, not the geocentric orientation. Therefore, the reference ellipsoid and normal ellipsoid were initially two different concepts.

5.2.2 Geometric Parameters of the Reference Ellipsoid and Their Correlations

The six commonly used geometric parameters in the Earth ellipsoid are as follows:

$$\left.\begin{array}{l}\text{Semimajor axis } a\\ \text{Semiminor axis } b\\ \text{Polar radius of curvature } c=\dfrac{a^2}{b}\\ \text{Flattening } f=\dfrac{a-b}{a}\\ \text{First eccentricity } e=\dfrac{\sqrt{a^2-b^2}}{a}\\ \text{Second eccentricity } e'=\dfrac{\sqrt{a^2-b^2}}{b}\end{array}\right\} \tag{5.9}$$

Out of the six parameters, given a single length parameter and another arbitrary parameter, the size and shape of the ellipsoid will be defined. In geodesy, a and f are customarily used to represent the geometric shape of the Earth ellipsoid.

Beijing Geodetic Coordinate System 1954 of China adopted the Krassowski Ellipsoid (see Table 4.1). Its parameters are as follows:

$a = 6{,}378{,}245.0000$ m
$b = 6{,}356{,}863.0188$ m
$c = 6{,}399{,}698.9018$ m
$f = 1/298.3 = 0.00335232986926$
$e^2 = 0.00669342162297$
$e'^2 = 0.00673852541468$

Xi'an Geodetic Coordinate System 1980 of China adopted the Geodetic Reference System 1975 (GRS75) reference ellipsoid recommended by International Union of Geodesy and Geophysics (IUGG) in 1975, also referred to as the IUGG-1975 Ellipsoid. Its geometric parameters are:

$a = 6{,}378{,}140.0000$ m
$b = 6{,}356{,}755.2882$ m
$c = 6{,}399{,}596.6520$ m
$f = 1/298.257 = 0.00335281317790$
$e^2 = 0.00669438499959$
$e'^2 = 0.00673950181947$

China Geodetic Coordinate System 2000 (CGCS2000) basically adopted the Geodetic Reference System 1980 (GRS80) recommended by IUGG in 1980 [*GM*

value of the Ellipsoid GRS80 is defined more precisely (see Nin et al. 2006)], and also referred to as the IUGG-1980 Ellipsoid. Its geometric parameters are:

$a = 6{,}378{,}137.0000$ m
$b = 6{,}356{,}752.3141$ m
$c = 6{,}399{,}593.6259$ m
$f = 1/298.257222101$
$e^2 = 0.00669438002290$
$e'^2 = 0.00673949677547$

By estimation, the values commonly used are:

$a \approx b \approx c \approx 6{,}400$ km
$f \approx 1/300$
$e^2 \approx e'^2 \approx 0.007 \approx 1/150$

a, b, c, f, e, and e' are the six geometric parameters often used in the Earth ellipsoid. To make it simpler to write and convenient to calculate, one can also introduce the auxiliary functions:

$$\left.\begin{aligned} W &= \sqrt{1 - e^2 \sin^2 B} \\ V &= \sqrt{1 + e'^2 \cos^2 B} \end{aligned}\right\}, \tag{5.10}$$

where B denotes the geodetic latitude, W is the first auxiliary function, and V is the second. Both are elliptic functions.

It is quite easy to obtain the relationship between the parameters from the definition of each parameter.

Relationship Between a and b

It follows from (5.9) that:

$$\frac{b^2}{a^2} = 1 - e^2, \frac{a^2}{b^2} = 1 + e'^2, \tag{5.11}$$

$$b = a\sqrt{1 - e^2}, \quad a = b\sqrt{1 + e'^2}. \tag{5.12}$$

Relationship Between e and e'

From (5.11) it can be found that:

$$\left(1 - e^2\right) \left(1 + e'^2\right) = 1$$

and hence:

$$
\begin{aligned}
&1-e^2=\frac{1}{1+e^{'2}}, 1+e^{'2}=\frac{1}{1-e^2},\\
&e^2=\frac{e^{'2}}{1+e^{'2}}, e^{'2}=\frac{e^2}{1-e^2},\\
&e=e^{'}\sqrt{1-e^2}, e^{'}=e\sqrt{1+e'^2}.
\end{aligned}
\tag{5.13}
$$

Relationship Between *a* and *c*

It follows from (5.9) and (5.11) that:

$$
a=c\sqrt{1-e^2},\ \ c=a\sqrt{1+e^{'2}}. \tag{5.14}
$$

Relationship Between *f* and *e*

From (5.9) and (5.11) one also obtains:

$$
f=1-\sqrt{1-e^2},\ \ e^2=2f-f^2. \tag{5.15}
$$

Relationship Between *W* and *V*

From (5.10) one gets:

$$
W^2=1-e^2\sin^2 B=1-\frac{e^{'2}}{1+e^{'2}}\left(1-\cos^2 B\right)=\frac{1+e^{'2}\cos^2 B}{1+e^{'2}}=\frac{V^2}{1+e^{'2}}.
$$

Hence:

$$
W=V\sqrt{1-e^2}, \text{and}\ \ V=W\sqrt{1+e^{'2}}. \tag{5.16}
$$

The above parameters, signs, and basic expressions of relationship will always be used hereafter in the derivations of formulae.

5.3 Relationship Between the Geodetic Coordinate System and the Geodetic Spatial Rectangular Coordinate System

5.3.1 Definitions of the Geodetic Coordinate System and the Geodetic Spatial Rectangular Coordinate System

The Geodetic Coordinate System is used to describe the geometric position of a point on the Earth's surface expressed by the geodetic longitude (L), geodetic latitude (B), and geodetic height (H).

As illustrated in Fig. 5.6, the geodetic longitude of point P_0 is the dihedral angle formed between the geodetic meridian plane NP_0S of an arbitrarily chosen point P_0 and the initial geodetic meridian plane NGS (meridian through the Greenwich Mean Observatory), denoted by L (initial letter of the German word *Länge*). Point P_0 is the projection of P on the ellipsoid along the surface normal. The geodetic longitude of point P is equal to that of its projection on the ellipsoid P_0. Longitudes can either be counted from the initial meridian plane eastward, ranging from 0° to 360°, or be measured eastward or westward from the Prime Meridian at Greenwich, ranging from 0° to 180° east or west, known as east longitude and west longitude, respectively. East longitudes are given positive values and west longitudes are given negative values in geodesy. Obviously, all points on the same meridian have the same longitude.

The geodetic latitude of point P is the angle formed from the equatorial plane to the ellipsoidal normal PK_p, denoted by B (initial letter of the German word *Breite*). The geodetic latitude of point P is equal to that of its projection on the ellipsoid P_0. Geodetic latitudes are measured southward or northward from the equator to poles, positively towards the north and negatively towards the south, ranging from 0° to 90°, known as south latitude and north latitude, respectively. Apparently, all points on the same parallel have the same latitude.

The geodetic height at a point P on the Earths' surface is the distance from the reference ellipsoid to the point in a direction normal to the ellipsoid, denoted by H (the distance H between P and P_0 in Fig. 5.6). Geodetic heights are measured from the ellipsoid, which are reckoned positive outward and negative inward.

The geodetic longitude L, geodetic latitude B, and geodetic height H constitute a three-dimensional geodetic coordinate system. These three coordinate values can uniquely specify the position of a point on the Earth's surface. If the point is on the surface of the ellipsoid, obviously $H = 0$, therefore, the position of a point on the ellipsoid can be uniquely determined by means of geodetic longitude L and geodetic latitude B, which is a two-dimensional geodetic coordinate system.

The direction of a curve on the ellipsoid is represented by the geodetic azimuth, which is the angle between the geodetic meridian and the geodetic line to the object

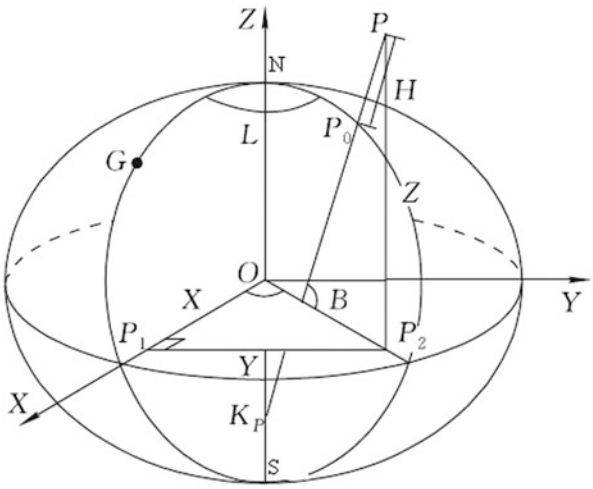


Fig. 5.6 Geodetic coordinate system and geodetic spatial rectangular coordinate system

observed, measured clockwise from the north meridian, ranging from 0° to 360°. This angle is denoted by A.

In Fig. 5.6, the origin of the geodetic spatial rectangular coordinate system (geodetic Cartesian coordinate system) is situated at the center O of the ellipsoid, and the line of intersection between the initial geodetic meridian and the equatorial plane is the X-axis. The Y-axis is perpendicular to the X-axis on the equatorial plane. The Z-axis is the spin axis of the ellipsoid, and the right-handed coordinate system O-XYZ is thus formed. The position of point P is represented by X, Y, and Z.

Definitions of geodetic and astronomical coordinate systems (cf. Sect. 2.1.3) are similar, although also different:

1. They have different reference surfaces and datum lines. The geodetic coordinate system uses the reference ellipsoid as its reference surface. Its datum line is the ellipsoidal normal, whereas the reference surface and datum line for the astronomical coordinate system are the geoid and the plumb line.
2. Geodetic coordinates are defined mathematically, whereas astronomical coordinates have physical meaning, influenced by the irregularity of the plumb line.
3. λ and φ are determined by theodolites directly, whereas L and B are calculated from observed quantities at some given point, including directions, distances, coordinate differences, etc.

5.3.2 Expressions of the Ellipsoidal Normal Length

As shown in Fig. 5.7, we establish a rectangular plane coordinate system, with axes x, y, and z, within a meridian plane or ellipse. Draw the normal PK_P through point P, and the angle between the normal PK_P and the x-axis is B. Through point P draw a line TP tangent to the meridian, and the angle between the tangent line TP and the

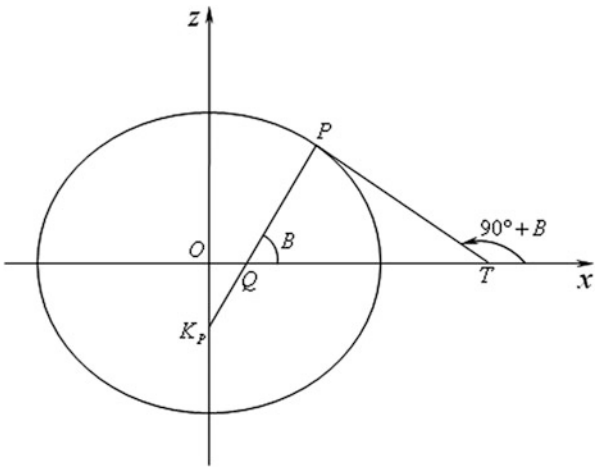


Fig. 5.7 The relationship between x, z, and B

x-axis is $90^\circ + B$. The slope of the tangent at P is equal to the first-order derivative of the curve at this point, and we have:

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \tan(90^\circ + B) = -\cot B. \tag{5.17}$$

So, the connection between x, z, and B can be found. Taking the derivative of the equation of meridian and substituting the above equation, the expressions of the two coordinate systems can be obtained.

The elliptic equation is:

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

Taking the derivative with respect to x gives:

$$\frac{2x}{a^2} + \frac{2z}{b^2} \cdot \frac{\mathrm{d}z}{\mathrm{d}x} = 0,$$

or

$$y = -\frac{b^2}{a^2} \frac{x}{\frac{\mathrm{d}z}{\mathrm{d}x}}.$$

With (5.17) and $b = a\sqrt{1-e^2}$, one obtains:

$$z = x\left(1 - e^2\right)\tan B.$$

Substituting the above equation into the elliptic equation, one obtains the parametric equations in terms of parameter B, as follows:

$$\left.\begin{aligned} x &= \frac{a}{\sqrt{1 - e^2\sin^2 B}}\cos B = \frac{a}{W}\cos B \\ z &= \frac{a(1 - e^2)}{\sqrt{1 - e^2\sin^2 B}}\sin B = \frac{a(1 - e^2)}{W}\sin B \end{aligned}\right\}.$$

In the above equations, $W = \sqrt{1 - e^2\sin^2 B}$ is introduced.
Let $PK_P = N$; from Fig. 5.7, we have:

$$\left.\begin{aligned} x &= PK_P\cos B = N\cos B \\ z &= PQ\sin B \end{aligned}\right\}.$$

Comparing the above two equations yields:

$$\left.\begin{aligned} PK_P &= N = \frac{a}{W} \\ PQ &= N(1 - e^2) \\ QK_P &= PK_P - PQ = Ne^2 \end{aligned}\right\}. \tag{5.18}$$

5.3.3 *Transformation Between Geodetic and Cartesian Coordinates*

As shown in Fig. 5.8, in the right triangle PK_pP_3, $K_pP_3 = (N + H)\cos B$, and $OP_2 = K_pP_3$, according to (5.18). Hence, in the right triangle OP_1P_2:

$$\begin{aligned} X &= OP_2\cos L = (N + H)\cos B\cos L, \\ Y &= OP_2\sin L = (N + H)\cos B\sin L. \end{aligned}$$

In the right triangle PQP_2:

$$PQ = N\left(1 - e^2\right) + H.$$

Hence:

$$Z = PP_2 = PQ\sin B = \left(N\left(1 - e^2\right) + H\right)\sin B.$$

The formulae to compute geodetic Cartesian coordinates using geodetic coordinates are given by:

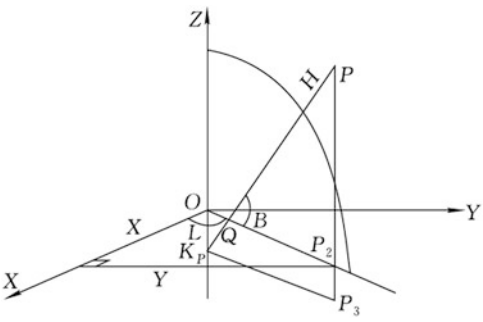


Fig. 5.8 The relationship between (L, B, H) and (X, Y, Z)

$$\left.\begin{aligned} X &= (N+H)\cos B\cos L \\ Y &= (N+H)\cos B\sin L \\ Z &= \left(N(1-e^2)+H\right)\sin B \end{aligned}\right\}. \tag{5.19}$$

The formulae used to calculate L, B, and H with X, Y, and Z is to be discussed here.

The first two equations in (5.19) divided by each other give:

$$\tan L = \frac{Y}{X},$$

then

$$L = \tan^{-1}\frac{Y}{X}. \tag{5.20}$$

Take out the meridian containing point P in Fig. 5.8, expressed by Fig. 5.9. Obviously, in Fig. 5.9, $OK_P = Ne^2 \sin B$, in the right triangle PK_PP_3, and we have:

$$\tan B = \frac{Z + N\cdot e^2 \sin B}{\sqrt{X^2+Y^2}}.$$

Substituting $N = \frac{a}{W}$ into the above equation and dividing the numerator and denominator by cos B yields:

$$\tan B = \frac{1}{\sqrt{X^2+Y^2}}\left(Z + \frac{ae^2\tan B}{\sqrt{1+\tan^2 B - e^2\tan^2 B}}\right). \tag{5.21}$$

B is unknown on the both sides of (5.21). Thus, tan$B$ has to be computed by iterations. Set the initial value tanB_0 of iteration on the right side:

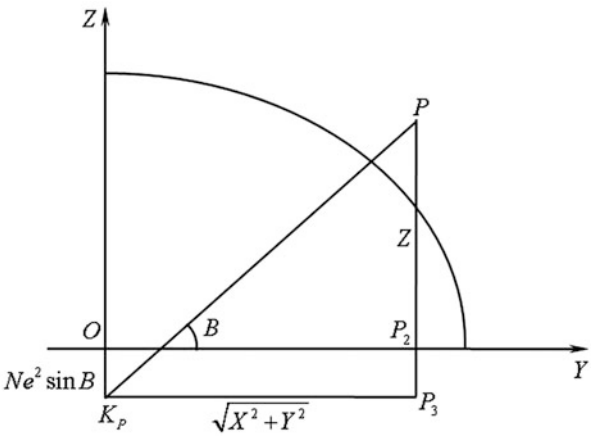


Fig. 5.9 The meridian that contains point P

$$\tan B_0 = \frac{Z}{\sqrt{X^2 + Y^2}}.$$

The tan$B$ on the left side (tanB_1) can be computed. Set tan$B$ = tanB_1; the iterations can be repeated until the difference between the two values of B is less than $0.0001''$ or the difference between the adjacent values of tan B is less than 5×10^{-10}.

From Fig. 5.9, in the right triangle PK_pP_3, it is apparent that:

$$\cos B = \frac{\sqrt{X^2 + Y^2}}{N + H}.$$

Hence, we get:

$$H = \frac{\sqrt{X^2 + Y^2}}{\cos B} - N. \tag{5.22}$$

Equations (5.20), (5.21), and (5.22) are the expressions for computing L, B, and H with X, Y, and Z.

Such transformation can be easily understood. The problem is that (5.21) needs iteration. There are many closed formulae for transformation between L, B, H, and X, Y, Z (see Featherstone and Claessens 2008; Vermeille 2004).

An example of transformation between L, B, H and X, Y, Z is shown in Table 5.1.

Table 5.1 Transformation between geodetic (L, B, H) and Cartesian (X, Y, Z) coordinates

$L, B, H \rightarrow X, Y, Z$		
Known data	Ellipsoidal parameters	Computational results (m)
$L = 77°11'22.333''$	Krassowski Ellipsoid	$X = 1{,}178{,}143.5316$
$B = 33°44'55.666''$		$Y = 5{,}181{,}238.3896$
$H = 5{,}555.660$ m		$Z = 3{,}526{,}461.5382$
	GRS75 Ellipsoid	$X = 1{,}178{,}124.3290$
		$Y = 5{,}181{,}153.9404$
		$Z = 3{,}526{,}400.6434$
	GRS80 Ellipsoid	$X = 1{,}178{,}123.7744$
		$Y = 5{,}181{,}151.5015$
		$Z = 3{,}526{,}399.0011$
$X, Y, Z \rightarrow L, B, H$		
Known data (m)	Ellipsoidal parameters	Computational results
$X = 1{,}177{,}888.777$	Krassowski Ellipsoid	$L = 77°\ 09'\ 27.2049''$
$Y = 5{,}166{,}777.888$		$B = 33°\ 57'\ 18.7484''$
$Z = 3{,}544{,}555.666$		$H = 3{,}878.5341$ m
	GRS75 Ellipsoid	$L = 77°\ 09'\ 27.2049''$
		$B = 33°\ 57'\ 18.8303''$
		$H = 3{,}984.3839$ m
	GRS80 Ellipsoid	$L = 77°\ 09'\ 27.2049''$
		$B = 33°\ 57'\ 18.8296''$
		$H = 3{,}987.3758$ m

5.4 Normal Section and Geodesic

5.4.1 Radius of Curvature of a Normal Section in an Arbitrary Direction

A plane containing the normal to the ellipsoid is called the normal section plane (Fig. 5.10). A normal section is created by intersecting the normal section plane with the surface of the ellipsoid, such as the meridian. An oblique section, on the other hand, is the intersection of the ellipsoid with any other plane that does not contain the surface normal, such as the parallel.

The normal section plays a vital role in geodetic computations. Observations of horizontal directions are usually referred to the direction of the plumb line, so if the plumb line coincides with the normal or coincides with the normal after corrections are applied, then the intersection of the ellipsoid with the vertical plane will be the normal section. In order to carry out geodetic computations on the surface of the ellipsoid, the properties of a normal section must be understood, and the radius of curvature of the normal section is one of the important concerns. At every point on the ellipsoid, infinitely many normal sections pass. In general, the radius of curvature varies with the direction of the normal sections. We will first derive the formula for radius of curvature of the normal section in an arbitrary direction, followed by that in special directions.

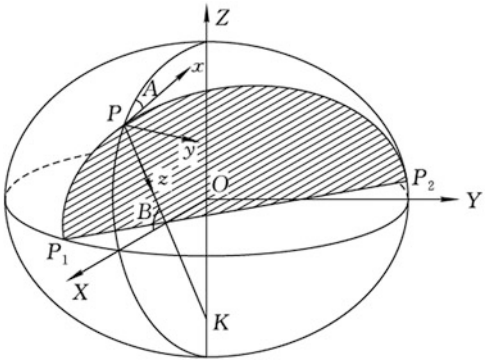


Fig. 5.10 Normal section in an arbitrary direction and the coordinate system

Overview

The intersection of a plane containing the normal to the ellipsoid with the ellipsoid surface will form the normal section. To solve simultaneous equations where one is the equation of an ellipsoid and the other is the equation of normal section plane will give the equation of the normal section, which is a plane curve. The radius of curvature of the normal section can be obtained according to the formula for radius of curvature of the plane curve.

In Fig. 5.10 we establish the Cartesian coordinate system *O*-*XYZ* with the origin at the center *O* of the ellipsoid. In this coordinate system, the equation of the ellipsoid is given by:

$$\frac{X^2}{a^2}+\frac{Y^2}{a^2}+\frac{Z^2}{b^2}=1. \tag{5.23}$$

Let *P* be a point on the ellipsoid. The curvature of the normal section at any point of the parallel passing through *P* on the rotational ellipsoid is the same in the same direction. In order to simplify derivations of formulae, let *P* be on the initial meridian plane. *PK* is the normal passing through point *P* and P_1PP_2 is the normal section passing through *P* in an arbitrary direction. Given the geodetic azimuth *A*, the equation of P_1PP_2 has to be found.

Because the normal section plane P_1PP_2 is intersecting the coordinate plane *O*-*XYZ*, one can imagine the complexity of its equations, making it inconvenient to solve the equation of the normal section. In the meantime, the normal section is represented by the space curve, so we cannot mechanically apply the formula for radius of curvature of a plane curve. To simplify the equation of the normal section plane, one needs to establish a new coordinate system. The curvature of a curve is a measure of how "curved" a curve is and it is independent of the choice of coordinate systems. In the newly established coordinate system, let a certain coordinate plane coincide with this normal section plane. Meanwhile, in order to compute the radius of curvature, assume that the origin of the newly established coordinate system *P*-*xyz* coincides with *P*, the *z*-axis coincides with the normal at point *P*, and the

x-axis coincides with the tangent line at point P. Together with the y-axis, the right-handed coordinate system is constituted (see Fig. 5.10). Obviously, in the new coordinate system, the equation of the normal section plane is $y = 0$. Put the equation of the normal section plane together with the equation of the ellipsoid to form simultaneous equations. The solutions to the simultaneous equations cannot be found unless the equation of the ellipsoid in the new coordinate system is obtained. In this case, the equation of the normal section will easily be solved, and the radius of curvature of the normal section can also be obtained.

To conclude, derivations of the formulae can be broken down into the following three steps:

1. Find the equation of the ellipsoid in the coordinate system P-xyz.
2. Find the equation of the normal section in any arbitrary direction.
3. Find the radius of curvature of the normal section in an arbitrary direction.

In formula derivations we need the formula for the coordinate transformation by rotating the systems and the formula for the radius of curvature of a plane curve, as given below.

We transform the right-handed Cartesian coordinate system by rotating the coordinate system through a counterclockwise angle θ_z about the Z-axis (θ_z is positive) according to right-hand rule; then:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old} = \mathbf{R}_Z(\theta_z) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old}.$$

$\mathbf{R}_Z$ is the rotation matrix. By the same token, we can obtain the transformation formulae of the coordinate system by rotating it about the X-axis and the Y-axis, and the rotation matrixes $\mathbf{R}_X$ and $\mathbf{R}_Y$ are given by:

$$\mathbf{R}_X(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix},$$

$$\mathbf{R}_Y(\theta_y) = \begin{bmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{bmatrix},$$

The rotation matrix satisfies the orthogonality condition (orthogonal matrix).

According to higher mathematics, the formula for the radius of curvature at point x_0 on the plane curve $y = f(x)$ can be written as:

$$R_{x_0} = \frac{\left[1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)_{x_0}^2\right]^{\frac{3}{2}}}{\left(\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\right)_{x_0}}.$$

Derivations of the Formula

Transformation Between Coordinate Systems *P-xyz* and *O-XYZ*

The equation of the ellipsoid in the coordinate system *P-xyz* cannot be computed unless the relationship of transformation between the coordinate systems *O-XYZ* and *P-xyz* is determined. As shown in Fig. 5.10, let P be a point on the initial meridian (the coordinate plane *XOZ)*; PK is the normal passing through P, $PK = N$, $PQ = N(1 - e^2)$, and the coordinate values of point P can be obtained from Fig. 5.10:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} N\cos B \\ 0 \\ N(1-e^2)\sin B \end{bmatrix}.$$

As shown in Fig. 5.10, in order to make point P coincide with point O, the origin of the coordinate system P should be translated to the point O. To make the coordinate plane xPz coincide with the meridian, the coordinate system should be rotated with a negative angle A around the z-axis. Similarly, for the z-axis, which is directed towards the normal, parallel to the minor axis of the ellipsoid, the coordinate system should be rotated by an angle of $90° + B$ about the y-axis. *P-xyz* is thereby transformed into *O-XYZ*, given by:

$$\begin{aligned} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} &= \mathbf{R}_y\left(90^\circ + B\right)\mathbf{R}_z(-A)\begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} N\cos B \\ 0 \\ N(1-e^2)\sin B \end{bmatrix} \\ &= \begin{bmatrix} -\sin B\cos A & \sin B\sin A & -\cos B \\ \sin A & \cos A & 0 \\ \cos B\cos A & -\cos B\sin A & -\sin B \end{bmatrix}\begin{bmatrix} x \\ y \\ y \end{bmatrix} \\ &\quad + \begin{bmatrix} N\cos B \\ 0 \\ N(1-e^2)\sin B \end{bmatrix}. \end{aligned} \tag{5.24}$$

Namely:

$$\left.\begin{aligned} X &= -(x\cos A - y\sin A)\sin B - z\cos B + N\cos B \\ Y &= x\sin A + y\cos A \\ Z &= (x\cos A - y\sin A)\cos B - z\sin B + N\left(1 - e^2\right)\sin B \end{aligned}\right\}. \tag{5.25}$$

Equation of the Ellipsoid in the Coordinate System *P-xyz*

Substituting (5.25) into (5.23) produces the equation of the ellipsoid in the coordinate system *P-xyz*.

To simplify the substitution, the (5.23) is rewritten in another form. Multiply both sides of the equation by a^2; with $\frac{a^2}{b^2} = 1 + e'^2$ and $a^2 = N^2W^2 = N^2(1 - e^2\sin^2 B)$, the equation of the ellipsoid in *O-XYZ* can be alternatively expressed as:

$$X^2 + Y^2 + Z^2 + e'^2 Z^2 - N^2\left(1 - e^2\sin^2 B\right) = 0. \tag{5.26}$$

Substituting (5.25) into the above equation, upon rearrangement, one obtains the equation of the ellipsoid in *P-xyz*:

$$x^2 + y^2 + z^2 - 2Nz + e'^2[(x\cos A - y\sin A)\cos B - z\sin B]^2 = 0. \tag{5.27}$$

Equation of the Normal Section in an Arbitrary Direction

In the coordinate system *P-xyz,* the normal section plane in any arbitrary direction coincides with the coordinate plane of *xPz*; thus the equation of an arbitrary normal section plane is given by:

$$y = 0.$$

Substituting $y = 0$ into (5.27) produces the equation of the normal section in any arbitrary direction:

$$x^2 + z^2 - 2Nz + e'^2(x\cos A\cos B - z\sin B)^2 = 0. \tag{5.28}$$

Radius of Curvature of the Normal Section in an Arbitrary Direction

Equation (5.28) shows that the normal section in an arbitrary direction is a plane curve. Its equation can be expressed as $z = f(x)$. Applying the formula for the radius of curvature of the plane curve produces:

$$R_A = \frac{\left[1 + \left(\frac{\mathrm{d}z}{\mathrm{d}x}\right)_P^2\right]^{\frac{3}{2}}}{\left(\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\right)_P}.$$

Since P is the coordinate origin, the z-axis and the x-axis are the normal and tangent to the normal section, respectively; obviously:

$$\left.\begin{array}{l} x_P = z_P = 0 \\ \left(\dfrac{\mathrm{d}z}{\mathrm{d}x}\right)_P = 0 \end{array}\right\}. \tag{5.29}$$

Hence, the radius can be expressed as:

$$R_A = \frac{1}{\left(\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\right)_P}.$$

It is obvious that solving R_A is made easier due to the establishment of the new coordinate system P-xyz. Here, $\left(\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\right)_P$ is the curvature of the normal section in an arbitrary direction at point P. Taking the derivative with respect to x in (5.28) repeatedly, and with (5.29), we obtain:

$$2 - 2N\left(\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\right)_P + 2e'^2 \cos^2 B \cos^2 A = 0.$$

Namely:

$$\left(\frac{\mathrm{d}^2 z}{\mathrm{d}x^2}\right)_P = \frac{1 + e'^2 \cos^2 B \cos^2 A}{N}.$$

Hence, the formula for the radius of curvature of the normal section in an arbitrary direction is expressed as:

$$R_A = \frac{N}{1 + e'^2 \cos^2 B \cos^2 A}. \tag{5.30}$$

This formula indicates that R_A is not only dependent on the latitude B of a given point but also on the azimuth A of the normal section. However, it is independent of the longitude L of the point. The point P is assumed on the initial meridian while deriving the formula. It is applicable everywhere in the world. Given the location of the point, its latitude B will be known. At this point, both N and cos$B$ are the specified constants. In this case, R_A varies only with the azimuth A of the normal section.

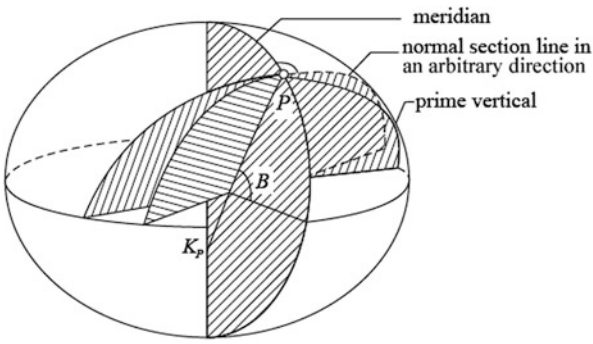


Fig. 5.11 Meridian and the prime vertical

5.4.2 *Radius of Curvature of the Meridian, Radius of Curvature in the Prime Vertical, and Mean Radius of Curvature*

Among the normal sections in all directions at point P on the ellipsoid, there exist two normal sections in special directions (cf. Fig. 5.11). One is the normal section with an azimuth of 0° (or 180°), i.e., the meridian. The other is the normal section with an azimuth of 90° (or 270°), i.e. the prime vertical. The radii of curvature in the meridian and the prime vertical are very frequently used in geodetic surveying computations.

Radius of Curvature in the Prime Vertical

Substituting $A = 90°$ into (5.30) yields:

$$R_{90} = N.$$

We assumed $PK_p = N$ in Fig. 5.11, because we know that the radius of curvature in the prime vertical is also defined as the normal to the ellipsoid terminating at the minor axis.

From (5.10) and (5.18) and with $a = c\sqrt{1-e^2}$, $W = V\sqrt{1-e^2}$, we get:

$$N = \frac{a}{W} = \frac{c}{V}. \tag{5.31}$$

Since $W = \sqrt{1 - e^2 \sin^2 B}$, and $V = \sqrt{1 + e'^2 \cos^2 B}$, it is obvious that N (radius of curvature) is a function of B (latitude) and increases with the increasing absolute value of B. The law of its changes is shown in Table 5.2.

Table 5.2 also indicates that when a point P on the ellipsoid moves from the equator along the meridian to the North Pole, the endpoint K of the radius of curvature in the prime vertical also moves relative to it. As shown in Figs. 5.12

Table 5.2 The law of changes in the radius of curvature in the prime vertical (N) with latitude (B)

B	N	Note
$B = 0^{\circ}$	$N_0 = \begin{cases} a \\ \dfrac{c}{\sqrt{1+e'^2}} \end{cases}$	At the equator, the prime vertical coincides with the equator. N is the equatorial radius of curvature
$0^{\circ} < B < 90^{\circ}$	$a < N < c$	N increases with the increasing latitude, its value ranging from a to c
$B = 90^{\circ}$	$N_{90} = \begin{cases} \dfrac{a}{\sqrt{1-e^2}} \\ c \end{cases}$	At the poles, the prime vertical and the meridian coincide with each other. N is the polar radius of curvature

and 5.13, its trace is marked by a straight line OK_N connecting the center of the ellipsoid O on the minor axis with the point K_N under it.

Radius of Curvature of the Meridian

Substituting $A = 0^{\circ}$ into (5.31) gives:

$$R_0 = M = \frac{N}{1 + e'^2 \cos^2 B} = \frac{N}{V^2},$$

And with (5.31) and $V^2 = \dfrac{W^2}{1 - e^2}$, we obtain:

$$M = \frac{a(1 - e^2)}{W^3} = \frac{c}{V^3}. \tag{5.32}$$

Similarly to N, M is also a function of B and increases with the increasing absolute value of latitude B. The law of its changes is detailed in Table 5.3.

Tables 5.2 and 5.3 show that c is the radius of curvature of the normal section at the pole. On the surface of the ellipsoid, the normal sections that pass through the poles are meridians. Hence, the polar radius of curvature c coincides with the meridional radius of curvature at the poles.

Table 5.3 also indicates that when a point P on the ellipsoid moves from the equator along the meridian to the North Pole, the endpoint K' of the meridional radius of curvature also moves relative to it. As shown in Fig. 5.13, its trajectory is marked by an asteroid $K'_0\, K' K_N$ close to the center of the ellipsoid.

In (5.30), when $A = 0^{\circ}$ (or 180°), the value of R_A is smallest; when $A = 90^{\circ}$ (or 270°), the value of R_A is largest. Hence, N and M are the maximum and minimum of R_A, which means that N is always greater than M except at the poles.

When A starts from due North 0° and increases eastward to 90°, R_A shifts progressively from M to N; when A starts from due East 90° and increases eastward to 180°, R_A shifts gradually from N to M. Therefore the radius of curvature R_A in any

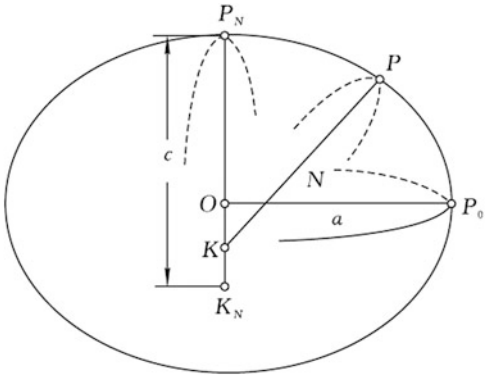


Fig. 5.12 Endpoint trace of N

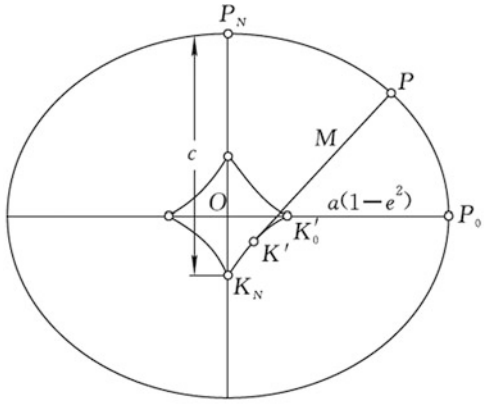


Fig. 5.13 Endpoint trace of M

Table 5.3 The law of changes in the radius of curvature of the meridian (M) with latitude (B)

B	M	Note
$B = 0^\circ$	$M_0 = \begin{cases} a(1-e^2) \\ \dfrac{c}{\left(1+e'^2\right)^{\frac{1}{2}}} \end{cases}$	At the equator, M is less than the equatorial radius a
$0^\circ < B < 90^\circ$	$a(1-e^2) < M < c$	M increases with the increasing latitude, its value ranging from $a(1-e^2)$ to c
$B = 90^\circ$	$M_{90} = \begin{cases} \dfrac{a}{\sqrt{1-e^2}} \\ c \end{cases}$	At the poles, M coincides with the polar radius of curvature c

arbitrary direction varies with the azimuth A with a cycle of 180°. Meanwhile, in (5.30), the only term related to the direction is $\cos^2 A$ and therefore the radii of curvature R_A at a given point in the direction of A, $(180^\circ - A)$, $(180^\circ + A)$, $(360^\circ - A)$ are the same. Hence, R_A changes over a cycle period of 180° and is symmetrical with respect to the meridian and the prime vertical.

Mean Radius of Curvature

The value of the mean radius of curvature, R_A, varies with direction, making the surveying computations difficult. Therefore, in practice the ellipsoid within a certain range will be considered as a sphere with an appropriate radius to meet the desired precision based on practical problems. It is reasonable to take the spherical radius as the mean radius of curvature R_A (averaging over all directions). The average of the radii of curvature of the normal sections in all directions at a point on the ellipsoid is the mean radius of curvature at this point, denoted by R.

To facilitate derivations, rearrange (5.30) as follows:

$$\begin{aligned} R_A &= \frac{N}{1+e'^2\cos^2 A\cos^2 B} = \frac{N}{\sin^2 A+\cos^2 A+e'^2\cos^2 A\cos^2 B} \\ &= \frac{MN}{M\sin^2 A+M\left(1+e'^2\cos^2 B\right)\cos^2 A} = \frac{MN}{N\cos^2 A+M\sin^2 A}. \end{aligned} \tag{5.33}$$

Since the variations in R_A with A are symmetrical with respect to the meridian and the prime vertical, one needs simply to find the average of R_A in one quadrant.

The average of the continuous function $y = f(x)$ on the closed interval $[a, b]$ is:

$$y_{average} = \frac{1}{b-a}\int_a^b f(x)\mathrm{d}x.$$

Hence, the average of R_A is:

$$R = \frac{1}{\frac{\pi}{2}-0}\int_0^{\frac{\pi}{2}} \frac{MN}{N\cos^2 A+M\sin^2 A}\mathrm{d}A = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \frac{\sqrt{MN}\sqrt{\frac{M}{N}}\frac{\mathrm{d}A}{\cos^2 A}}{1+\left(\sqrt{\frac{M}{N}}\tan A\right)^2}.$$

Set $t = \sqrt{\frac{M}{N}}\tan A$; thus $\mathrm{d}t = \sqrt{\frac{M}{N}}\frac{\mathrm{d}A}{\cos^2 A}$. The above equation upon substitution gives:

$$R = \frac{2}{\pi}\sqrt{MN}\int_0^{\infty}\frac{\mathrm{d}t}{1+t^2} = \sqrt{MN}. \tag{5.34}$$

This equation shows that the mean radius of curvature at a point on the ellipsoid equals the geometric mean of the radii of curvature in the meridian and the prime vertical at this point.

Substituting the expressions of M (5.32) and N (5.31) into (5.34) gives the formula for computing R:

$$R = \frac{a\sqrt{1-e^2}}{W^2} = \frac{c}{V^2}. \tag{5.35}$$

M, N, and R at a point on the ellipsoid are all measured inward along the normal at this point to the surface. Their lengths are generally different from each other. Comparing (5.31), (5.32), and (5.35), one can obtain their relations as:

$$N > R > M \quad (0 \leq B < 90^{\circ}).$$

When the point is at the poles, we have:

$$N_{90} = R_{90} = M_{90} = c \quad (B = 90^{\circ}).$$

Hence, c is the radius of curvature at the poles.

5.4.3 *Length of a Meridian Arc and Length of a Parallel Arc*

When geodetic computations are performed on the surface of the ellipsoid, such as computation of the Gauss projection, the formulae for the length of a meridian arc and the length of a parallel arc are needed. Derivations of the formulae are thereby given next.

Formula for Length of a Meridian Arc

As shown in Fig. 5.14, let P_1 and P_2 be two points along the meridian at latitudes B_1 and B_2, respectively; compute the meridian arc length X between the two points P_1 and P_2.

If the meridian is a circular arc, the length of the arc is the product of the radian measure of the central angle subtended by the arc times the radius of the circle. However, the meridian is an elliptical arc. Its arc length must be calculated using integration. Take a short arc segment on the meridian, i.e., the arc element (differential of arc) $PP' = \mathrm{d}X$. The central angle subtended by the arc (difference in latitude) is denoted by $\mathrm{d}B$. Let the point P be at the latitude B, and the latitude of point P' will be $B + \mathrm{d}B$. Let the meridional radius of curvature at the point P be M, and arc element $\mathrm{d}X$ can be considered the circular arc of radius M; thus:

$$\mathrm{d}X = M\mathrm{d}B. \tag{5.36}$$

To compute the arc length X between P_1 and P_2 is to find the integral of $\mathrm{d}X$ between B_1 and B_2, namely:

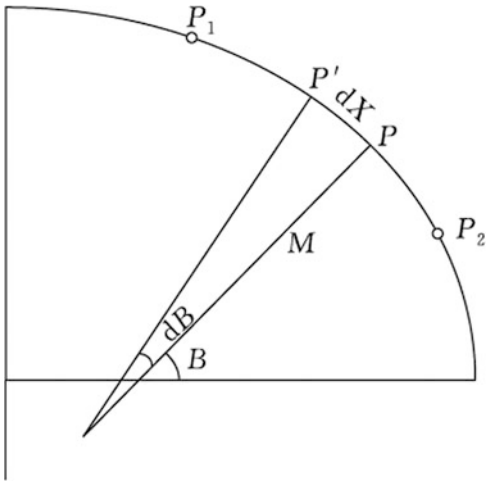


Fig. 5.14 Length of a meridian arc

$$X = \int_{B_2}^{B_1} \mathrm{d}X = \int_{B_2}^{B_1} M\mathrm{d}B,$$

with (5.32) we have:

$$X = a\left(1 - e^2\right)\int_{B_1}^{B_2} \left(1 - e^2 \sin^2 B\right)^{-\frac{3}{2}} \mathrm{d}B. \tag{5.37}$$

This function is an elliptic integral, which cannot be integrated directly; one needs to apply the binomial theorem:

$$(1-x)^n = 1 - \frac{n}{1!}x + \frac{n(n-1)}{2!}x^2 - \frac{n(n-1)(n-2)}{3!}x^3 + \cdots \quad (x < 1).$$

Expanding the function into a series:

$$\left(1 - e^2 \sin^2 B\right)^{-\frac{3}{2}} = 1 + \frac{3}{2}e^2 \sin^2 B + \frac{15}{8}e^4 \sin^4 B + \frac{35}{16}e^6 \sin^6 B + \cdots.$$

To simplify the integration process, one needs to change the power function of sine ($\sin^n B$) into the multiple-angle function of cosine ($\cos nB$), namely:

$$\begin{aligned}
\sin^2 B &= \frac{1}{2} - \frac{1}{2}\cos 2B, \\
\sin^4 B &= \frac{3}{8} - \frac{1}{2}\cos 2B + \frac{1}{8}\cos 4B, \\
\sin^6 B &= \frac{5}{16} - \frac{15}{32}\cos 2B + \frac{3}{16}\cos 4B - \frac{1}{32}\cos 6B, \\
&\cdots
\end{aligned}$$

Hence, it follows that:

$$\begin{aligned}
&(1 - e^2 \sin^2 B)^{-\frac{3}{2}} = 1 + \frac{3}{4}e^2 - \frac{3}{4}e^2\cos 2B + \frac{45}{64}e^4 - \frac{15}{16}e^4\cos 2B \\
&+ \frac{15}{64}e^4\cos 4B + \frac{175}{256}e^6 - \frac{525}{512}e^6\cos 2B + \frac{105}{256}e^6\cos 4B - \frac{35}{512}e^6\cos 6B \\
&+\cdots \qquad \cdots \qquad \cdots \\
&= A' - B'\cos 2B + C'\cos 4B - D'\cos 6B + E'\cos 8B - F'\cos 10B,
\end{aligned} \tag{5.38}$$

where the coefficients in the equation are:

$$\left.\begin{aligned}
A' &= 1 + \frac{3}{4}e^2 + \frac{45}{64}e^4 + \frac{175}{256}e^6 + \frac{11025}{16384}e^8 + \frac{43659}{65536}e^{10}\cdots \\
B' &= \frac{3}{4}e^2 + \frac{15}{16}e^4 + \frac{525}{512}e^6 + \frac{2205}{2048}e^8 + \frac{72765}{65536}e^{10}\cdots \\
C' &= \frac{15}{64}e^4 + \frac{105}{256}e^6 + \frac{2205}{4096}e^8 + \frac{10395}{16384}e^{10}\cdots \\
D' &= \frac{35}{512}e^6 + \frac{315}{2048}e^8 + \frac{31185}{131072}e^{10}\cdots \\
E' &= \frac{315}{16384}e^8 + \frac{3465}{65536}e^{10}\cdots \\
F' &= \frac{639}{131072}e^{10}\cdots
\end{aligned}\right\}.$$

These coefficients are all constants for the defined ellipsoid.
Substituting (5.38) into (5.37), integration yields:

$$X = a(1-e^2)\left[A'\frac{B_2-B_1}{\rho} - \frac{B'}{2}(\sin 2B_2 - \sin 2B_1) + \frac{C'}{4}(\sin 4B_2 - \sin 4B_1)\right.$$
$$\left. - \frac{D'}{6}(\sin 6B_2 - \sin 6B_1) + \frac{E'}{8}(\sin 8B_2 - \sin 8B_1) - \frac{F'}{10}(\sin 10B_2 - \sin 10B_1) + \cdots\right] \tag{5.39}$$

This is the general formula for the meridian arc length. In the ensuing terms like sin8B and sin10B, the greatest value of the term sin8B is only 0.0003 m. Thus, whether to leave it out is determined by the desired accuracy of computations.

In practical use, the formula for the meridian arc length from the equator to the point at latitude B is usually applied. In this case, inserting $B = 0$, $B_2 = B$ into (5.39) yields:

$$X = a(1-e^2)\left[A'\frac{B}{\rho} - \frac{B'}{2}\sin 2B + \frac{C'}{4}\sin 4B - \frac{D'}{6}\sin 6B + \frac{E'}{8}\sin 8B - \frac{F'}{10}\sin 10B + \cdots\right] \tag{5.40}$$

The formula (5.40) gives the meridian arc length from the equator to a given point along the meridian.

Substituting the defining parameters of the Krassowski Ellipsoid adopted by the Beijing Geodetic Coordinate System 1954 into the above equation yields:

$$X = 111134.8611B^{\circ} - 16036.4803\sin 2B + 16.8281\sin 4B - 0.0220\sin 6B + \cdots \tag{5.41}$$

Likewise, for the GRS75 Ellipsoid adopted by the Xi'an Goedetic Coordinate System 1980, we have:

$$X = 111133.0047B^{\circ} - 16038.5282\sin 2B + 16.8326\sin 4B - 0.0220\sin 6B + \cdots \tag{5.42}$$

For the GRS 80 Ellipsoid adopted by CGCS2000, we get:

$$X = 111132.95254700B^{\circ} - 16038.508741268\sin 2B + 16.832613326622\sin 4B - 0.021984374201268\sin 6B + 3.1141625291648\times 10^{-5}\sin 8B. \tag{5.43}$$

B° in (5.41), (5.42), and (5.43) denotes the geodetic latitude in degrees. X is measured in meters. If the length of the meridian arc X is known, and then the corresponding geodetic latitude B can be solved by the iteration method.

If we put $B = \frac{\pi}{2}$, placed in (5.43) the length of a quadrant of the meridian Q is 10,001,965 m.

That is, the length of a quadrant of a meridian is approximately 10,000 km, and then the circumference of the Earth is approximately 40,000 km. The length of a "meter" was originally defined as a ten-millionth of this length. In 1793, this length became the standard in France. However, it was later found that the first prototype meter bar was short by a fifth of a millimeter because of miscalculation of the flattening of the Earth. The polar circumference of the Earth is therefore shown to be about 8,000 m more than 40 million meters.

When the arc length is short (e.g., $X < 45$ km, calculations are accurate to 0.001 m), the meridian can be considered a circular arc. The radius of the circle is the meridional radius of curvature M_m at the mean latitude $B_m = \frac{1}{2}(B_1 + B_2)$ of the arc. The central angle equals the difference in latitude between the two endpoints, namely $\Delta B = B_2 - B_1$. Its computational formula is given by:

$$X = M_m \frac{\Delta B}{\rho},$$

with $\rho'' = 206264.80624710''$, or $\rho^{\circ} = 57.29577951308^{\circ}$.

Formula for Length of a Parallel Arc

A parallel (circle of latitude) is a circle, so the arc length along a parallel is the length of the circular arc of its subtended angle at the center (difference in longitude).

In Fig. 5.15, let P_1 and P_2 be two points on the parallel. Their latitude is B, difference in longitude is l, the radius of the parallel is r, and $P_1K = N$ is the radius of curvature in the prime vertical; we have:

$$r = N \cos B. \tag{5.44}$$

Hence, the formula for the length of a parallel arc can be written as:

$$S = r\frac{l}{\rho} = N \cos B \frac{l}{\rho}. \tag{5.45}$$

Variations in the Unit Meridian and Parallel Arc Lengths with Latitude

The formulae for the arc elements of the meridian and the parallel are given by:

$$\left.\begin{aligned} \mathrm{d}X &= M\mathrm{d}B \\ \mathrm{d}S &= r\mathrm{d}L \end{aligned}\right\}.$$

The meridian radius of curvature M increases gradually with increasing latitude, whereas the radius of the parallel r decreases sharply with increasing latitude. Hence, the meridian arc length of the unit latitude difference increases slowly

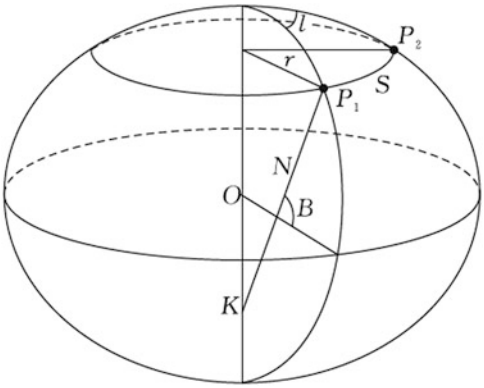


Fig. 5.15 Length of a parallel arc

with increasing latitude, short in the south and long in the north. The parallel arc length of the unit longitude difference decreases sharply with increasing latitude, long in the south and short in the north, as depicted in Fig. 5.16.

Table 5.4 lists the arc lengths at different latitudes for a quantitative grasp of the meridian and parallel arc lengths.

Table 5.4 outlines the variations in lengths of a meridian arc and a parallel arc with latitude. The length of 1° of latitude averages 110 km; 1′ (one minute) of latitude covers approximately 1.8 km and 1″ (one second) is about 30 m. The length of a parallel arc is almost equivalent to the length of the meridian arc near the equator, but they increasingly diverge from each other with increasing latitude.

M and R are more or less the same. In some approximate calculations, the Earth can be considered as a sphere. The relationships between the arc length on the sphere and its subtended angle at the center are as follows:

1° arc length $\approx$ 110 km
1′ arc length $\approx$ 1.8 km
1″ arc length $\approx$ 30 m

and

1 km $\approx$ 30″ arc length
1 m $\approx$ 0.03″ arc length
1 cm $\approx$ 0.0003″ arc length

We know that 1 nautical mile $=$ 1.852 km, which corresponds approximately to the value of a minute of arc along a meridian. In fact, the nautical mile was defined as the average length along the meridian arc represented by one minute of latitude.

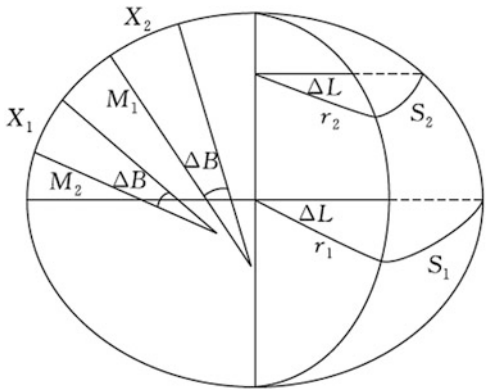


Fig. 5.16 Variations in meridian and parallel arc lengths

Table 5.4 Variations in meridian and parallel arc lengths with latitude B (GRS80 Ellipsoid)

	Length of a meridian arc (m)			Length of a parallel arc (m)		
B	$\Delta B = 1^{\circ}$	$\Delta B = 1'$	$\Delta B = 1''$	$l = 1^{\circ}$	$l = 1'$	$l = 1''$
0°	110,574	1,842.91	30.715	111 321	1,855.36	30.923
15°	110,653	1,844.15	30.736	107 552	1,792.54	29.876
30°	110,861	1,847.54	30.792	96 488	1,608.13	26.802
45°	111,141	1,852.20	30.870	78 848	1,341.14	21.902
60°	111,421	1,856.87	30.948	55 801	930.02	15.500
75°	111,623	1,860.30	31.005	28 902	481.71	8.028
90°	111,694	1,861.57	31.026	0.000	0.000	0.000

Area of the Trapezoidal Map Sheet

We will next discuss the formula for calculating the area of the trapezoidal map sheet as an application instance of the formulae for lengths of the meridian and parallel arcs.

Topographic maps are bounded by lines of latitude and longitude, which means that the surface of the ellipsoid is divided up into a series of map sheets according to certain differences in longitude and latitude. As shown in Fig. 5.17, BA and CD are lines of longitude, and BC and AD are lines of latitude. The coordinates of point A are (B_1, L_1) and the coordinates of point C are (B_2, L_2). An area element dP within the trapezoidal map sheet (quadrilateral), and with sides rdl and MdB has:

$$\mathrm{d}P = MN\cos B\mathrm{d}B\mathrm{d}L.$$

Hence, with $MN = \dfrac{b^2}{(1 - e^2\sin^2 B)^2}$, obtained from (5.35) and (5.12), the area P of the quadrilateral $ABCD$ is:

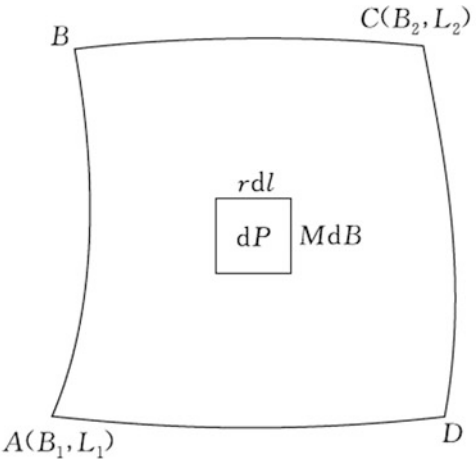


Fig. 5.17 Trapezoidal map sheet

$$P = \int_{L_1}^{L_2}\int_{B_1}^{B_2} MN\cos B\mathrm{d}B\mathrm{d}L = lb^2\int_{B_1}^{B_2}\frac{\cos B}{\left(1-e^2\sin^2 B\right)^2}\mathrm{d}B, \tag{5.46}$$

where $l = L_2 - L_1$. Expanding the above function according to binomial theorem and integrating gives:

$$P=\frac{l^\circ}{90^\circ}\pi b^2\left[A'\sin\frac{1}{2}(B_2-B_1)\cos B_m - B'\sin\frac{3}{2}(B_2-B_1)\cos 3B_m\right.$$
$$\left.+C'\sin\frac{5}{2}(B_2-B_1)\cos 5B_m - D'\sin\frac{7}{2}(B_2-B_1)\cos 7B_m + E'\sin\frac{9}{2}(B_2-B_1)\cos 9B_m\right]. \tag{5.47}$$

In the equation, $B_m = \frac{1}{2}(B_1+B_2)$. The coefficients are:

$$\left.\begin{aligned}
A' &= 1+\frac{1}{2}e^2+\frac{3}{8}e^4+\frac{5}{16}e^6+\frac{35}{128}e^8\\
B' &= \frac{1}{6}e^2+\frac{3}{16}e^4+\frac{3}{16}e^6+\frac{35}{192}e^8\\
C' &= \frac{3}{80}e^4+\frac{1}{16}e^6+\frac{5}{64}e^8\\
D' &= \frac{1}{112}e^6+\frac{5}{156}e^8\\
E' &= \frac{5}{2304}e^8
\end{aligned}\right\}.$$

Assume a trapezoidal map sheet at 1:1,000,000 scale, the difference in the south and north bounding latitude is $B_2 - B_1 = 4^\circ$ and the difference in the east and west bounding longitude is $l = 6^\circ$; substituting into the above equation yields the area of this trapezoid:

$$P = \frac{\pi b^2}{15}\left[A' \sin 2^\circ \cos B_m - B' \sin 6^\circ \cos 3B_m + C' \sin 10^\circ \cos 5B_m - D' \sin 14^\circ \cos 7B_m + E' \sin 18^\circ \cos 9B_m\right].$$

In (5.47), if we put $l = 360^\circ$, $B_1 = 0^\circ$, $B_2 = 90^\circ$, and multiply by 2, one obtains the formula for total surface area of the ellipsoid:

$$\begin{aligned}\Sigma &= 4\pi b^2\left[A' + B' + C' + D' + E' + \cdots\cdots\right] \\ &= 4\pi b^2\left[1 + \frac{2}{3}e^2 + \frac{3}{5}e^4 + \frac{4}{7}e^6 + \frac{5}{9}e^8 + \cdots\cdots\right] .\end{aligned} \tag{5.48}$$

The total area of GRS80 Ellipsoid is approximately 510,065,597 km^2. It thus follows that the total surface area of the Earth is about 5.1×10^8 km^2.

5.4.4 Reciprocal Normal Sections

Let us first consider the normal sections formed by reciprocal observations from two points on the surface of the ellipsoid. As shown in Fig. 5.18, A and B are two points on the surface of the ellipsoid. Their normals to the ellipsoid AK_a and BK_b and the corresponding plumb lines are assumed to be coincident at the two points. If A and B are the observation points, then the vertical plane will be the normal section plane. The instrument is set up at A and then aimed at point B defining the normal section by marking a point a; that is, AaB. The normal section AaB from point A to B is formed by the intersection of a plane containing the normal at point A and that passes through point B (the vertical plane AK_aB) with the surface of the ellipsoid. Similarly, the instrument set up at B and A is sighted defining the normal section marking a point b; that is, BbA. The normal section BbA from point B to A is formed by the intersection of the sighting plane BK_bA with the surface of the ellipsoid. The normal sections originating at points A and B, and terminating at B and A, respectively, namely AaB and BbA, generally do not coincide. They are termed the reciprocal normal sections between points A and B.

We can imagine that if the surface normals at two points A and B are in the same plane, then the two vertical planes at one point and passing through the other point in the reciprocal observations coincide. Only one normal section is formed. Otherwise, the two vertical planes do not coincide and at any of the points there are two

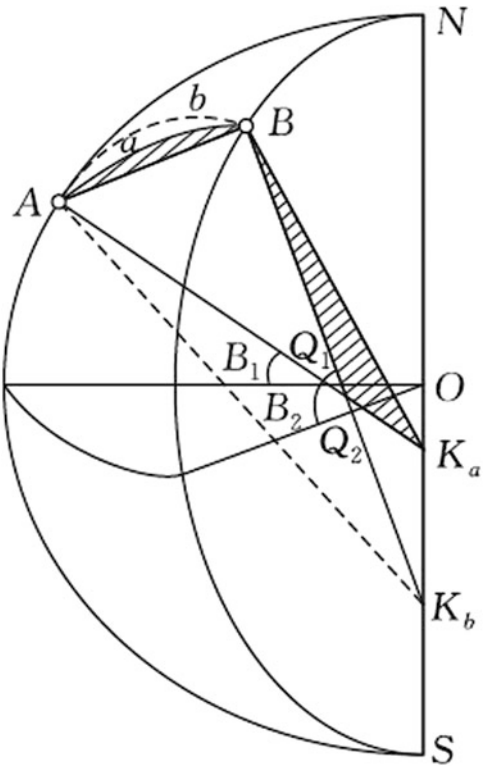


Fig. 5.18 Reciprocal normal section

normal sections. It can be seen that the reciprocal normal sections occur just because the normals at points A and B are not in the same plane.

Note that for the two points A and B on the surface of the ellipsoid with different longitudes and latitudes, the normal AK_a at the point A will not lie in the same plane as the normal BK_b at the point B.

As we know, if the normals AK_a and BK_b lie in the same plane, the two straight lines are either parallel to or intersecting each other. As shown in Fig. 5.18, A and B are not on the same meridian and thus have different longitudes. Generally, the angles that the normals make with the minor axis vary, so the two normals are not parallel to each other. Because the minor axis is the line of intersection of two meridian planes, if the normals on the two meridian planes intersect each other, the point of intersection will be at the minor axis.

Let the latitudes of the two points be B_1 and B_2; then the normals AK_a and BK_b intersect the equatorial plane at point Q_1 and Q_2, and from the Fig. 5.18 one obtains:

$$\begin{aligned} OK_a &= Q_1 K_a \sin B_1 \\ OK_b &= Q_2 K_b \sin B_2 \end{aligned}.$$

With the length of the normal on the lower side of the equator $QK = Ne^2$, we have:

$$\left.\begin{aligned} OK_a &= N_1 e^2 \sin B_1 \\ OK_b &= N_2 e^2 \sin B_2 \end{aligned}\right\}.$$

From the above equations, if $B_1 \neq B_2$, then $OK_a \neq OK_b$, so K_a and K_b do not coincide, which indicates that AK_a and BK_b do not lie in the same plane. When both points A and B are situated on the same meridian or on the same parallel, the normal section and reverse normal section will be coincident, which is a special case.

The above equations also show that when $B_2 > B_1$, then $OK_b > OK_a$. Figure 5.18 indicates that K_a is above K_b. The two normal section planes AK_aB and BK_bA intersect at the chord AB. The normal sections AaB and BbA are created by intersecting the planes containing the normals to the ellipsoid (normal section planes) with the surface of the ellipsoid. BbA is on the upper side, while AaB is on the lower side. It follows that the normal section from the point with higher latitude to the point with lower latitude is on the upper side, whereas that from the point with lower latitude to the point with higher latitude is on the lower side. We term AaB the normal section of point A, while BbA is the reverse normal section of A. According to the above law, we can draw the relationship of position between the normal section and reverse normal section from point A to B in different quadrants, as illustrated in Fig. 5.19.

Reciprocal normal sections usually do not coincide, except for the two cases where the two points are situated on the same meridian or at the same latitude. The angle Δ of the normal section and the reverse normal section between two points (Fig. 5.20) can be up to $0.004''$ when their distance apart S is 25 km or even a few hundredths of a second when $S = 50$ km (directly proportional to the square of the distance) in first-order triangulation. It must be taken into account in computation of first-order triangulation.

The existence of reciprocal normal sections brings inconvenience for geodetic computations. Let A, B, and C be three points on the surface of the ellipsoid. Their longitudes are $L_C > L_B > L_A$, and latitudes $B_B > B_C > B_A$. In reciprocal triangular observations, the case in Fig. 5.20 will occur where the three angles A, B, C formed by the normal sections cannot constitute a triangle, which is to say, the reciprocal normal sections have caused fracture of the geometric figure.

Obviously we cannot base our computations on a fractured figure but instead choose a unique curve between the two points to replace the reciprocal normal sections. There are many kinds of single curves between two points on the ellipsoid. However, the curve between the two points should be unique and possess distinct geometric properties (such as the shortest line between two points on the ellipsoid),

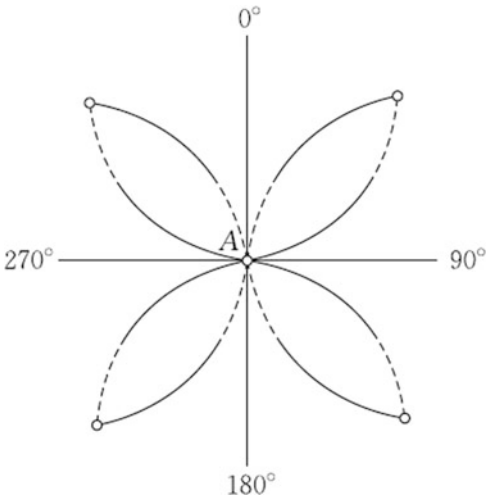


Fig. 5.19 Positions of reciprocal normal sections

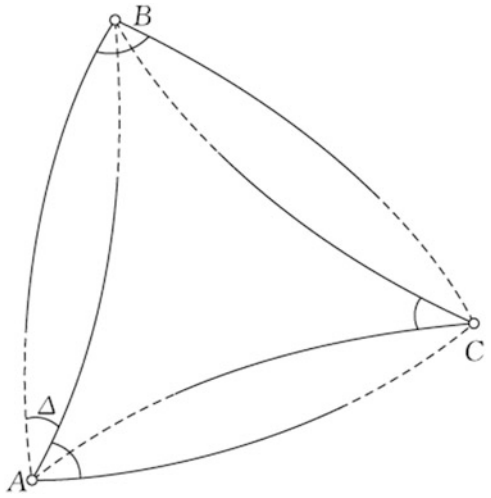


Fig. 5.20 Figure formed by reciprocal normal sections

making it convenient to perform survey computations on the surface of the ellipsoid.

5.4.5 The Geodesic

Definition of the Geodesic

The geodesic is defined as a curve on which the adjacent two arc elements of each point lie in the same normal section plane of this point. We can use Fig. 5.21 to make this definition clear. Let AB be a geodesic on the surface, P be an arbitrary point on the geodesic, $\mathrm{d}S_1$ and $\mathrm{d}S_2$ be the adjacent arc elements of point P, and PK be the normal to the surface at point P. $\mathrm{d}S_1$ and $\mathrm{d}S_2$ are both arc elements, i.e., the points P_1 and P_2 are infinitely close to point P, so we can use chords PP_1 and PP_2 to substitute $\mathrm{d}S_1$ and $\mathrm{d}S_2$. Consequently, $\mathrm{d}S_1$ lies in the normal section plane PKP_1 and $\mathrm{d}S_2$ lies in the normal section plane PKP_2. According to the definition of geodesic, the above two normal section planes will be in the same normal section plane at point P. To put it another way, the infinitely near points, P_1, and P_2 are all in the same normal section plane at point P. If every point on the same curve shares this property, then this curve will be the geodesic.

The geodesic can also be defined as a curve on a surface where at each point of the curve the principal normal of the curve coincides with the normal to the surface of the ellipsoid at this point. It is quite convenient to determine whether a curve on the surface of the ellipsoid is a geodesic according to this definition. We will elaborate on this below.

For space curves, the line perpendicular to the tangent line to a curve at the point is called the normal to the curve at that point. Thus, the space curve at a given point has a bunch of normals. The aggregation of the normals forms a plane called the normal plane. The principal normal to a curve at a point is a special normal on the normal plane that points towards the concave side of the curve. This definition is consistent with the previous one since taking an arbitrary arc element means that the concave side of the curve at this point, namely the principal normal to the curve at this point, is determined. Moreover, the normal section is determined by the normal to the surface. The arc element is required to lie in the normal section, and then the principal normal to a curve coincides with the normal to the surface.

Properties of the Geodesic

The Geodesic Is the Line of Shortest Distance Between Two Points on the Ellipsoid

In Fig. 5.21, with the projection of the adjacent two arc elements of point P on the geodesic orthogonally onto a plane tangent to the ellipsoid at this point, one obtains $P_1'PP_2'$. Since the three points are on the same normal section plane, $P_1'PP_2'$ is a straight-line element. The shortest path between two points on any plane is a straight line. The orthogonal projection of the adjacent two arc elements of any point on the geodesic are the straight-line elements. Hence, the geodesic is the

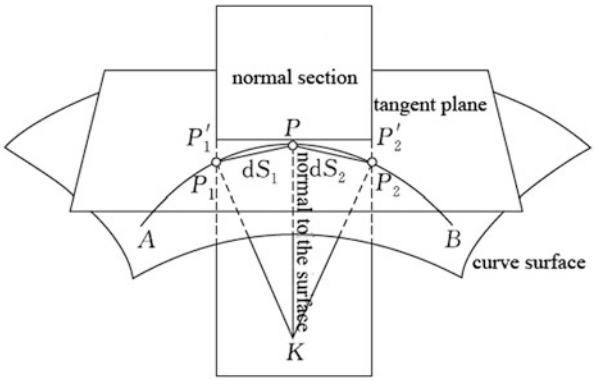


Fig. 5.21 Definition of geodesic (two sides of the two triangles in the middle of the figure, left-side dS_1 and right-side dS_2)

shortest route between two points. However, the projection of the arc element of other curves like the arc element of an oblique curve onto the tangent plane will be a curve element without fail. The parallel is an oblique curve. In Fig. 5.22, PP_1 andPP_2 are the two adjacent arc elements of point P on the parallel. Its orthographic projection onto the tangent plane T at point P is the curve element $P_1^{'}PP_2^{'}$.

The Geodesic Is the Connection of the Arc Elements of Numerous Normal Sections

The adjacent two arc elements of a point on the geodesic are on the same normal section plane, and hence they can be considered the arc elements of two normal sections with their orientations 180° apart at this point. Therefore, the geodesic is the connection of the arc elements of these normal sections at each point. If we draw a straight-line traverse on the ellipsoid, as shown in Fig. 5.23, and let the deflection angle be 180°, then the sides are so short that the normal section and reverse normal section coincide with each other. As shown in Fig. 5.23, *ab* and *ba* coincide and become $\overline{ab}$. This short-side straight-line traverse is the geodesic.

The normal sections on the ellipsoid are not geodesics except for the meridian and equator. Note that the normal section is a curve lying in a plane that contains the normal at one point and passes through the other point, whereas the geodesic is any normal section that passes through every point along the curve.

The Geodesic Is a Curve of Double Curvature on the Surface of the Ellipsoid

Geodesics are on the surface of the ellipsoid. The bending of the ellipsoid causes the longitudinal bending of geodesics, which is represented by the curvature at each point. Since each point along the geodesic has a different longitude and latitude, the

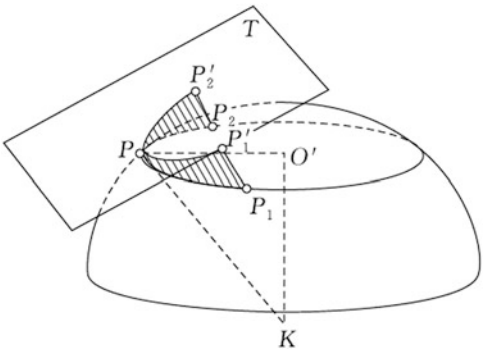


Fig. 5.22 Orthographic projection of the arc element on a parallel

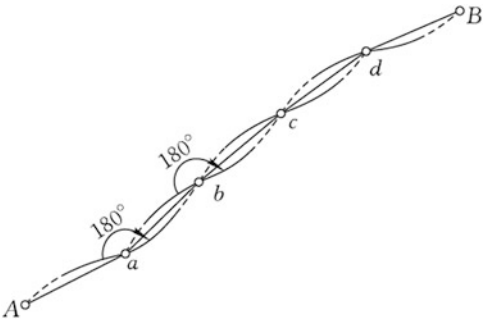


Fig. 5.23 Traversing the geodesic

normals to the surface at each point do not intersect, and normal section planes do not coincide. Thus, lateral bending of geodesics arises, which is represented by the torsion at each point (see Fig. 5.24). Therefore, the geodesics on the ellipsoid are curves with both curvature and torsion except for the meridian and equator.

If an elastic band is stretched between two points on an absolutely smooth ellipsoid, then this tightly stretched rubber band is the geodesic between two points. The direction of the compressive stress exerted by the rubber band on the ellipsoid at each point is the principal normal to the curve. The direction of the anchorage force provided by the ellipsoid at this point is the normal to the surface. When the rubber band goes slack, the principal normal to the curve coincides with the normal to the surface. Therefore, due to the existence of elasticity, the rubber band will always represent the shortest path between two points.

The Geodesic Lies Between Two Reciprocal Normal Sections

In general, on the surface of the ellipsoid, the geodesic lies between two reciprocal normal sections and is close to the direct normal section. In addition, it divides the angle between the reciprocal normal sections at a ratio of approximately 2 to 1, i.e. $\mu : \gamma = 2 : 1$, as shown in Fig. 5.25. The value of γ can range from $0.001''$ to $0.002''$ in first-order triangulation when S is around 35 km. Corrections need to be

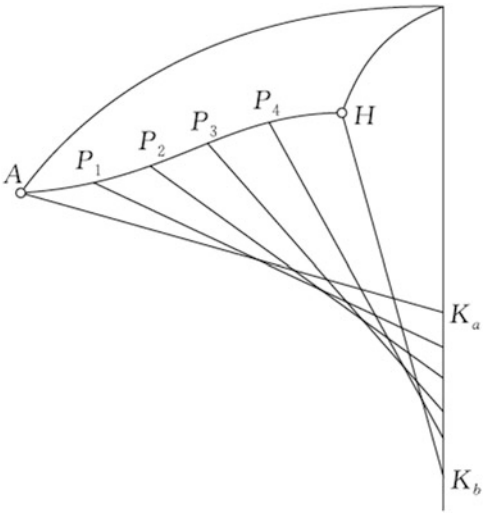


Fig. 5.24 The shape of geodesics

taken into consideration in computation (called correction from normal section to geodesic). The difference in length between the normal section and the geodesic is slight. If the geodetic latitude at a given point $B = 0°$, the geodetic azimuth of the side $A = 45°$, and the length of the side $S = 100$ km, then the difference in the length between the normal section and the geodesic $\Delta S = 0.000001$ mm, which can be shown to be negligible in practical cases.

On the meridian and equator, the geodesic coincides with the reciprocal normal sections and both the geodesic and the reciprocal normal sections coincide with the meridian and the equator. On the parallel circle, although the normal section and the reverse normal section coincide, the geodesic, the normal sections, and the parallel do not coincide.

Differential Equations of the Geodesic

Differential equations of geodesics are the differential expressions of the relationships between the geodesic distance, geodetic longitude, latitude, and azimuth.

As shown in Fig. 5.26, let P be an arbitrary point on the geodesic. Its longitude is L, latitude is B, and geodetic azimuth is A. Let PP_1 be the arc element of the geodesic dS. From point P to point P_1, its longitude changes into L+dL, latitude is B+dB, and azimuth is A+dA.

From Fig. 5.26, the arc element of the meridian $P'P_1 = M\mathrm{d}B$, and the arc element of the parallel $PP' = r\mathrm{d}L = N\cos B\mathrm{d}L$. $PP'P_1$ is a right ellipsoidal triangle. Since it is infinitesimal, the right ellipsoidal triangle can be considered a right-angled plane triangle. Hence, one obtains:

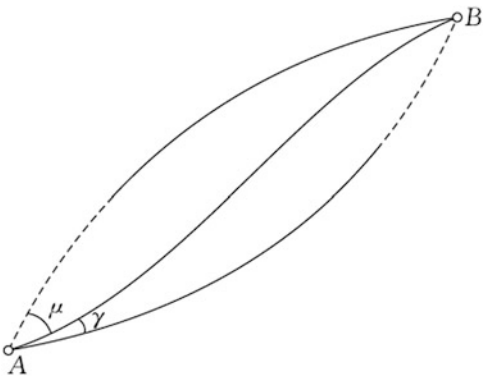


Fig. 5.25 Positional relationship of geodesic and reciprocal normal sections

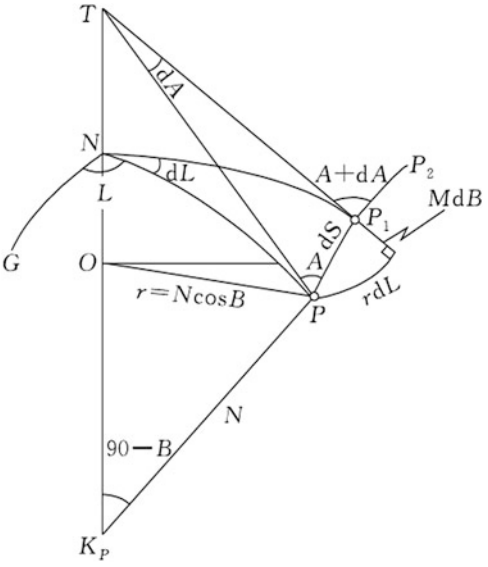


Fig. 5.26 Differential relationships of geodesics

$$M\mathrm{d}B = \mathrm{d}S\cos A.$$

We see that:

$$\mathrm{d}B = \frac{\cos A}{M}\mathrm{d}S, \tag{5.49}$$

and $r\mathrm{d}L = \mathrm{d}S\ \sin\ A$.

$$\text{So } \mathrm{d}L = \frac{\sin A}{r}\mathrm{d}S = \frac{\sin A}{N}\ \sec B\mathrm{d}S. \tag{5.50}$$

The above two differential equations are obtained under the condition that the triangle is infinitesimal. It applies to any curves on the surface of the ellipsoid, of course including geodesics. However, it is not exclusively applicable to geodesics. In order to derive the differential equations exclusive to geodesics, the definition of geodesics should be taken into consideration.

In Fig. 5.26, PP_1 is the arc element of point P_1 on the geodesic. Let P_1P_2 be another arc element adjacent to P_1. In accordance with the definition of geodesics, PP_1 and P_1P_2 are on the same normal section plane at point P_1. Hence, PP_1P_2 is the arc element of the normal section at point P_1. Its orthogonal projection onto the tangent plane at point P_1 is a straight-line element.

We draw a tangent plane to a surface at point P_1, and a line tangent to the meridian through P_1 and P. Because P and P_1 are infinitely near points, the two tangent lines can be considered to meet at point T on the extension of the minor axis. The plane defined by P_1T and PT can also be considered the tangent plane to the surface at point P_1. Hence, the arc element of the geodesic PP_1P_2 can be considered a straight line on the tangent plane. $\angle\, TP_1P_2 = A + \mathrm{d}A$. It is an exterior angle of the plane triangle TPP_1. One can obtain:

$$A + \mathrm{d}A = A + \angle P_1TP, \quad \text{with} \quad \mathrm{d}A = \angle P_1TP.$$

Since $\mathrm{d}S$ is the arc element, P' is infinitely close to P_1; then, P' can be considered on the plane tangent to the surface at point P_1. Hence, with the minor sector TPP', one obtains:

$$\mathrm{d}A = \frac{r\mathrm{d}L}{PT} = \frac{N\cos B\mathrm{d}L}{PT}.$$

It follows from the right triangle K_PPT that $PT = N\cot B$, and substitution into the above equation yields:

$$\mathrm{d}A = \sin B\mathrm{d}L.$$

Inserting (5.50) into the above equation gives:

$$\mathrm{d}A = \frac{\sin A}{N}\tan B\mathrm{d}S. \tag{5.51}$$

Equations (5.49), (5.50), and (5.51) are generally referred to as the three differential equations of geodesics. They are the precondition for computing geodetic coordinates on the surface of the ellipsoid. If an increment is employed to replace the differential, one can obtain the formulae for approximately computing the differences in geodetic latitude, longitude, and azimuth as follows:

$$\left.\begin{aligned}\Delta B &= \frac{S}{M}\cos A \\ \Delta L &= \frac{S}{N}\sin A \sec B \\ \Delta A &= \frac{S}{N}\sin A \tan B\end{aligned}\right\}. \tag{5.52}$$

Clairaut's Equation for the Geodesic

Integrating the differential equations of geodesics gives Clairaut's equation. It provides the basis for solving the geodetic problems over long distance.

From (5.51) we can get:

$$\mathrm{d}A = \frac{\sin A \sin B}{N \cos B}\mathrm{d}S,$$

with (5.49), $\mathrm{d}S = \frac{M}{\cos A}\mathrm{d}B$, we obtain:

$$\mathrm{d}A = \frac{\sin A}{\cos A} \cdot \frac{M \sin B \mathrm{d}B}{N \cos B},$$

where, as shown in Fig. 5.27, P is a point on the surface of the ellipsoid, PP' is the arc element of the meridian, and $PP' = M\mathrm{dB}$. The difference in the radius of the parallel at point P and P' *is* $\mathrm{d}r$. When P' moves to P, the latitude increases and the radius of parallel decreases. Let $PP'P''$ be a small right-angled plane triangle; we get:

$$M \sin B \mathrm{d}B = -\mathrm{d}r,$$

with $r = N \cos B$, substituting into the above equation gives:

$$\mathrm{d}A = -\frac{\sin A}{\cos A} \cdot \frac{\mathrm{d}r}{r}.$$

We obtain:

$$r \cos A \mathrm{d}A + \sin A \mathrm{d}r = 0,$$

$$\cot A \mathrm{d}A = -\frac{\mathrm{d}r}{r}.$$

After integration, it can be rewritten as:

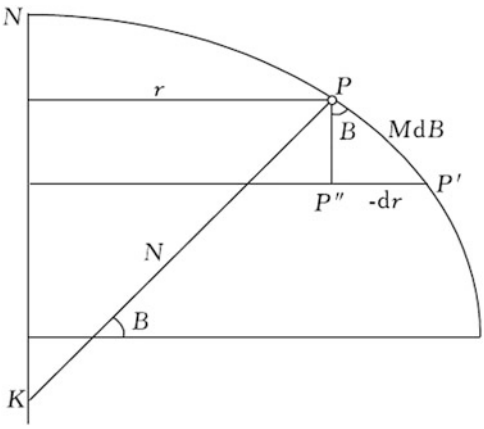


Fig. 5.27 Computation of dr using meridian arc element

$$r \sin A = C. \tag{5.53}$$

This is known as Clairaut's equation for the geodesic (e.g., as described in Thomas and Featherstone 2005), which shows that the product of the radius of the parallel and the sine of the geodesic azimuth at any point along the geodesic is a constant.

5.4.6 Solution of Ellipsoidal Triangles

After establishing the corresponding relationships between the Earth's surface and the surface of the ellipsoid, the terrestrial control network will become the control network formed by geodesics on the ellipsoid. The unobserved sides and angles of the control network must be computed using ellipsoidal triangles. Since the curvatures at each point on the ellipsoid are different, it is rather complicated to solve triangles on the ellipsoid. However, the flattening of the Earth ellipsoid is very small, and the ellipsoidal triangle formed by the three sides in the geodetic control network is usually quite small; therefore, it may be possible to solve the ellipsoidal triangle as a spherical triangle. Research has shown that when the sides of the triangle are less than 200 km, it is completely viable to approximate the ellipsoidal triangle as a spherical triangle, and the spherical radius is the mean radius of curvature at the mean latitude of the three vertices of the ellipsoidal triangle. (The sides of the spherical triangle are equal in length to the corresponding sides of the ellipsoidal triangle, while the difference in the corresponding angles of the two triangles is less than $0.001''$.)

The sine formula below is applied to solve the spherical triangle:

$$\frac{\sin\frac{a}{R}}{\sin A}=\frac{\sin\frac{b}{R}}{\sin B}=\frac{\sin\frac{c}{R}}{\sin C}.$$

The lengths of sides in the above equation are expressed by angles. The formula can only be applied if the known lengths of sides are divided by the spherical radius to get the angle at the center of the sphere. The calculated lengths of sides that are expressed in degrees should be converted to lengths again because in practical cases the sides are always represented by lengths. It often appears inconvenient to convert. In the meantime, there also exist some round-off errors in computation that will adversely influence the accuracy of computation.

We will attempt to find a simpler way to solve the triangle. Here, Legendre's theorem is a simple and convenient way to solve the spherical triangle. The Legendre method, in nature, is to solve the spherical triangle as the plane triangle that has the same corresponding sides as the spherical triangle. The spherical angles are required to make some simple changes.

As shown in Fig. 5.28, let $A_0B_0C_0$ be a spherical triangle, a, b, and c are its three sides, and ε'' denotes the spherical excess. We draw a planar triangle $A_1B_1C_1$ with the same side lengths of a,b, and c. When the side lengths are not long, the three internal angles in the two triangles can be proved to have the following relations:

$$\left.\begin{aligned}A_1&=A_0-\varepsilon''/3\\B_1&=B_0-\varepsilon''/3\\C_1&=C_0-\varepsilon''/3\end{aligned}\right\},\tag{5.54}$$

where $\varepsilon''=\frac{\Delta}{R^2}\rho''$, Δ is the area of the plane triangle, and R denotes the radius of the sphere.

The above equation is Legendre's theorem to solve spherical triangles. It shows that one-third of the spherical excess of the given spherical triangle $A_0B_0C_0$ subtracted from each angle of the triangle gives the plane triangle $A_1B_1C_1$ whose sides are equal in length to the corresponding sides of the spherical triangle. The side lengths computed according to the formulae for the plane triangle will be the side lengths of the spherical triangle.

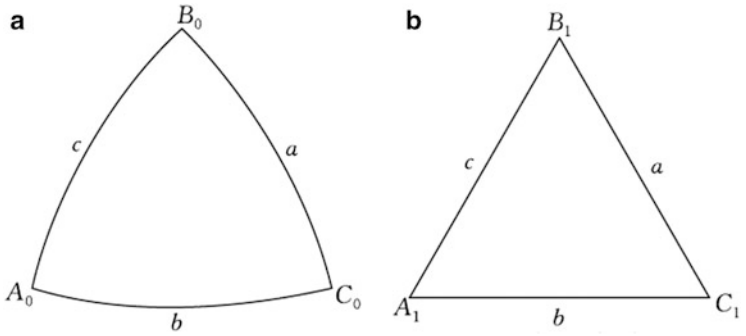


Fig. 5.28 The spherical triangle (**a**) and plane triangle (**b**) whose corresponding sides are equal in length

5.5 Relationship Between Terrestrial Elements of Triangulateration and the Corresponding Ellipsoidal Elements

5.5.1 Significance of and Requirements for Reduction of Terrestrial Triangulateration Elements to the Ellipsoid

Conventional geodesy determines the horizontal coordinates L and B of an Earth's surface point and the height of a point above the Earth's surface as two separate issues. In order to find L and B of a surface point, one needs to project the geodetic control network actually established on the physical surface of the Earth to the reference ellipsoid adopted, i.e., to reduce the geodetic observations to the reference ellipsoid (see, e.g., Torge and Müller 2012). That is to say, the sides measured on the physical surface of the Earth between points must be reduced to the geodesic arc length on the surface of the reference ellipsoid and the observed values of horizontal directions and astronomical azimuths should be reduced to the geodesic directions and geodetic azimuths (Fig. 5.29). Then, all the calculations concerning the geodetic control network will be carried out on the surface of the ellipsoid as a two-dimensional problem. After adjustment computations, one will find the longitude L and latitude B, which can be transformed into the plane coordinates x, y by applying the specified mathematical relations. Adjustment computations of the large-scale astro-geodetic network are usually completed as such.

For small-scale geodetic control networks, it will be inconvenient to carry out adjustment computations on the ellipsoid since the mathematical properties of the ellipsoid is far more complicated than that of the plane. Hence, again we can project the elements of the geodetic control network on the surface of the ellipsoid onto the plane and then carry out computations on the planar surface, (see Chap. 6).

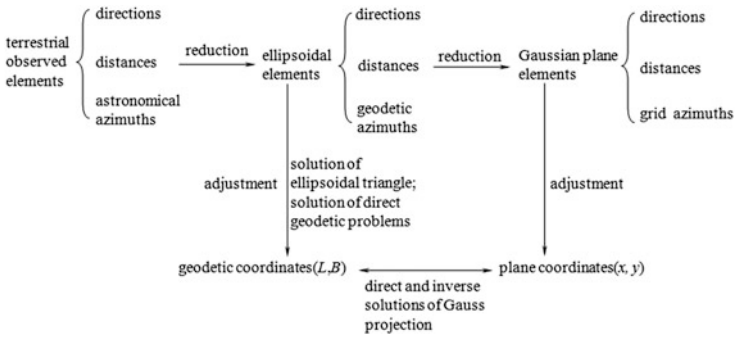


Fig. 5.29 Diagram for solving horizontal coordinates using observed values

Reduction of the elements on the Earth's surface to the surface of the ellipsoid means that appropriate corrections are applied to the terrestrial observation elements. In this case, the terrestrial observation elements will be transformed into the corresponding elements on the ellipsoid, making geodetic computations on the ellipsoid possible. The degree of accuracy of reductions should not jeopardize the accuracy of field observations.

There are three fundamental requirements for reducing the elements of terrestrial triangulateration to the ellipsoid surface.

1. The ellipsoidal normal is referred to as the reference line. The direction of the plumb line at the observation point usually does not coincide with the corresponding normal to the ellipsoid. The angle between the plumb line and the line normal to the ellipsoid at the same point is called deflection of the vertical. The ellipsoidal normal is made the datum, which means that the values observed referring to the plumb line will be converted to the values that are referred to the ellipsoidal normal. In this case, the effect of deflection of the vertical will be removed.
2. The ellipsoid is taken as the reference surface. We project the point from the physical surface of the Earth onto the ellipsoid along the ellipsoidal normal and reduce the observed values at a certain height above the Earth's surface to those on the surface of the ellipsoid. In this case, the effect of the ellipsoidal height will be eliminated.
3. The geodesic is taken as the connection between two points on the ellipsoid. Reduce the observed values in the direction of the normal section to the observed values in the direction of the geodesic, and the effect of the normal section-geodesic separation will be removed.

Through the above reductions, the terrestrial triangulateration network will be reduced to the ellipsoid, as shown in Fig. 5.30. It thus follows that the reduction essentially achieves transformations between elements on the Earth's surface and elements on the surface of the ellipsoid.

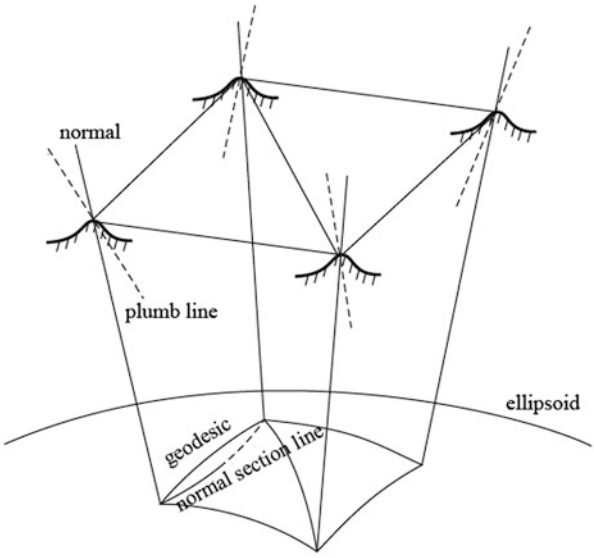


Fig. 5.30 Reduction of terrestrial control network to ellipsoid

Reductions of terrestrial triangulateration elements to the ellipsoid include the reductions of horizontal directions, observed zenith distances, terrestrial distances, as well as the reductions of the astronomical azimuths, etc. In addition, computations of the deflection of the vertical, as an essential element in reduction, are also one of the concerns in this section.

It should be noted that, although GPS surveying techniques can directly determine the geodetic coordinates of the position of a point, in applications where directions, distances, or azimuths on different reference surfaces are needed, we should still perform the corresponding calculations as shown in Fig. 5.29.

5.5.2 Reduction of Horizontal Directions to the Ellipsoid

Reducing horizontal directions to the ellipsoid requires correction for deflection of the vertical, correction for skew normals, and correction from normal section to geodesic. These three corrections are customarily known as the "three corrections for horizontal directions".

Correction for Deflection of the Vertical

The Earth's surface point is projected onto the ellipsoid along the ellipsoidal normal. The direction of the plumb line at the observation point does not coincide with the corresponding ellipsoidal normal, which will exert effects on the values of

observed directions. This effect is termed the correction for deflection of the vertical, denoted by δ_1.

This correction serves to solve the problem of relations between spatial angles. One effective means of solving such a problem is to create an auxiliary sphere and to denote the spatial angles by the arc length of the auxiliary sphere; namely, to reduce the spatial angles to angles on the spherical surface. Then, the problem can be solved by employing the methods of solving spherical triangles.

As shown in Fig. 5.31, we create an auxiliary sphere with the observation point A as the center. Then we draw a normal passing through point A that intersects the auxiliary sphere at Z, which is called the geodetic zenith; the plumb line passes through A, which intersects the auxiliary sphere at Z_1, which is called the astronomical zenith. The deflection of the vertical μ is the angle between the ellipsoidal normal and the plumb line at the point. We draw a line through point A parallel to the minor axis of the ellipsoid, intersecting the auxiliary sphere at P, known as the north pole of the auxiliary sphere. $ZP\theta$ is the meridian on the auxiliary sphere, which is the line of intersection between the ellipsoidal meridian plane and the auxiliary sphere. M is the projection of the terrestrial target point m on the surface of the ellipsoid. With the north pole and the meridian introduced, the azimuth in the Am direction can be expressed on the auxiliary sphere since spatial angles do not change by translating a straight line or plane. The deflection of the vertical μ on the auxiliary sphere is a short arc length, so it can be decomposed into two perpendicular components ξ and η, which are known as the components of the deflection of the vertical μ in the meridian and in the prime vertical, respectively, as seen in Fig. 5.31:

$$\xi = \mu \cos\theta, \quad \eta = \mu \sin\theta.$$

From Fig. 5.31, if M lies in the vertical plane ZZ_1O, whether the observed directions are referred to the ellipsoidal normal or the plumb line, the vertical plane is the same. In this instance, there will be no vertical deflection correction. Hence, we refer to AO as the reference direction (zero direction of the circle). In Fig. 5.31, if measured with respect to the plumb line AZ_1, the circle reading at the target point m will be OR_1; measured with respect to the normal AZ, the circle reading at the target point m will be OR. Hence, the effect of the deflection of the vertical on the horizontal directions will be $\delta_1 = (R - R_1)$.

R_1MR is a right spherical triangle. By applying the law of sines, we have:

$$\sin(R_1R) = \sin(-\delta_1) = \sin(90 - z_1)\sin q,$$

where z_1 is the observed zenith distance in the M direction. Applying the sine theorem to the triangle MZZ_1 we get:

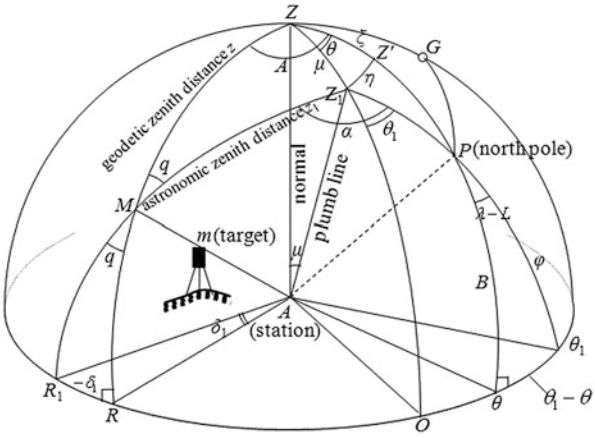


Fig. 5.31 Correction for the deflection of the vertical

$$\sin q = \sin \mu \frac{\sin R}{\sin z_1}.$$

Substituting this equation into the above equation results in:

$$\sin(-\delta_1) = \cos z_1 \sin \mu \frac{\sin R}{\sin z_1}.$$

Since both δ and μ are small quantities, we can write:

$$-\delta_1 = \mu \sin R \cot z_1,$$

with $R = A - \theta$ (A denotes the geodetic azimuth from point A to point m); then:

$$\begin{aligned} -\delta_1 &= \mu \sin(A - \theta) \cot z_1 \\ &= \mu(\sin A \cos\theta - \cos A \sin\theta) \cot z_1. \end{aligned}$$

Given $\xi = \mu \cos\theta$, $\eta = \mu \sin\theta$, we obtain:

$$\begin{aligned} \delta_1 &= -(\xi \sin A - \eta \cos A) \cot z_1 \\ &= -(\xi \sin A - \eta \cos A) \tan \alpha_1, \end{aligned} \tag{5.55}$$

where α_1 is the vertical angle between the line of sight and the horizontal. It can be seen that the correction for deflection of the vertical is primarily concerned with the deflection of the vertical at the observation point and the zenith distance (vertical angle) at the target point.

In Fig. 5.31 the horizontal circles perpendicular to the plumb line and the ellipsoidal normal do not coincide, and the angle between them is μ. However, they are considered to coincide because μ is a small quantity and its effect on the horizontal circle reading can be neglected.

In general, μ can range from a few seconds to over $10''$. The vertical angle is very close to 0° in the first- and second-order triangulation. In plain areas, the vertical angle is usually about $\pm 30'$, and in mountainous areas, it can presumably amount to $\pm 3^\circ$. Hence, the value of δ_1 is usually a few tenths of a second. Corrections for deflection of the vertical need to be applied to the first- and second-order triangulations. If the deflection of the vertical and the vertical angle are both quite large, correction for deflection of the vertical should be taken into consideration even in the third- or fourth-order triangulations.

On the following occasions, correction for deflection of the vertical equals zero:

1. The plumb line coincides with the ellipsoidal normal, i.e., $\mu = 0$, then $\delta_1 = 0$
2. The target point is in the plane ZZ_1O, i.e., $A = \theta$, then $\delta_1 = 0$
3. The target point is in the horizontal plane, i.e., $Z_1 = 90^\circ$, then $\delta_1 = 0$

Correction for Skew Normals

After the correction δ_1 is applied, the direction value will be the direction of the normal section Ab' in Fig. 5.32. Here, the height of the observation point A has no effect on the value of horizontal directions. To simplify, we set A on the ellipsoid. In accordance with the requirements for reductions, the projection of the target point B on the surface of the ellipsoid should be b rather than b'. Hence, the angular difference between these two normal sections Ab' and Ab is δ_2, known as the skew normal correction. Obviously, this correction is due entirely to the effect of the height of the target point B on the reduced direction value.

In Fig. 5.32, the ellipsoidal triangle Abb' is considered a plane triangle. By applying the sine theorem, it gives:

$$\delta_2 = bb' \frac{\sin A_1}{S} \rho''.$$

This serves to show that bb' needs to be computed first in order to obtain δ_2. In the triangles Bbb' and BRK_a, we see that:

$$bb' = H_2\theta,$$
$$\theta = \frac{K_a R}{BR} \approx \frac{K_a R}{N_2}.$$

The "2" denotes the corresponding value at point B and the subscript "1" in the following equations denotes the corresponding value at point A. From Fig. 5.32:

$$K_a R = K_a K_b \cos B_2,$$

and

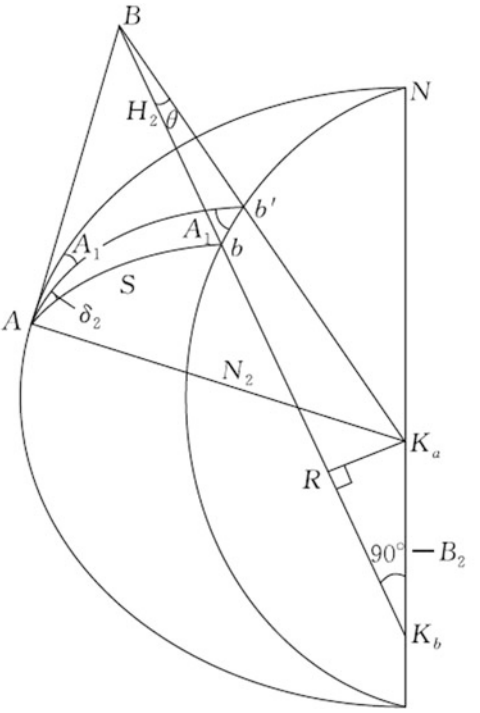


Fig. 5.32 Correction for skew normals

$$K_aK_b = OK_b - OK_a = N_2e^2 \sin B_2 - N_1e^2 \sin B_1 \approx N_2e^2(\sin B_2 - \sin B_1),$$

With the differential equation of geodesics $B_2 \approx B_1 + \dfrac{S \cdot \cos A_1}{M_1}$, one obtains:

$$K_aK_b = N_2e^2 \cos B_1 \frac{S \cos A_1}{M_1}.$$

Inserting back into the above equation and rearranging gives:

$$\delta_2 = \frac{e^2H_2\rho''}{2M_2} \cos^2 B_2 \sin 2A_1, \tag{5.56}$$

where B_2 and M_2 are the geodetic latitude and meridian radius of curvature of the target point, A_1 is the geodetic azimuth of the observed direction, and H_2 is the height of the target (actual target point in observation) relative to the ellipsoid, consisting of three parts:

$$H_2 = H_{2N} + \zeta_2 + a_2,$$

where H_{2N} is the normal height of the monument center of target point, ζ_2 is the

height anomaly of the target point, and a_2 is the height of the structure at the target point.

It follows from (5.56) that the correction for skew normals is dependent primarily on the height of the target point.

Let $B_2 = 45°$, $A_1 = 45°$, and $\rho''/M_2 = 1/30$; when $H_2 = 200$ m, $\delta_2 = 0.01''$; when $H_2 = 1{,}000$ m, $\delta_2 = 0.05''$. It can be seen that correction for the skew normals should be taken into consideration in the first- and second-order triangulation and in the third- and fourth-order triangulation at high altitudes.

On the following occasions, the correction for skew normals is zero:

1. The target point is on the ellipsoid, i.e., $H_2 = 0$; then $\delta_2 = 0$
2. The target point and the observation point are at the same longitude or latitude, i.e., $A_1 = 0°, 90°, 180°, 270°$, then $\delta_2 = 0$

Correction from Normal Section to Geodesic

Connections between two points on the ellipsoid are referred to the geodesic. Therefore, in Fig. 5.32 the normal section direction *Ab* should be converted to the corresponding geodesic direction, known as the normal section to geodesic correction, denoted by δ_3 (as illustrated in Fig. 5.33).

The formula for correction from normal section to geodesic is:

$$\delta_3'' = -\frac{e^2 S^2 \rho''}{12 N_1^2} \cos^2 B_1 \sin 2A_1. \tag{5.57}$$

Obviously, the correction from normal section to geodesic is chiefly concerned with the distance from the place of observation to the target point.

Set $B_1 = 45°$, $A_1 = 45°$; when $S = 30$ km, $\delta_3 = -0.001''$; when $S = 60$ km, $\delta_3 = -0.005''$. This serves to show that correction from normal section to geodesic should generally be applied to the first-order triangulation. However, second- or lower order triangulation can be carried out irrespective of this correction.

When $A_1 = 0°, 90°, 180°$, and $270°$, $\delta_3 = 0$, i.e., points *A* and *B* are on the same meridian or approximately on the same parallel, and the correction from normal section to geodesic is zero.

Computation of the Three Corrections for Horizontal Directions

The three corrections for horizontal directions are theoretical problems in classical geodesy, indicating the relations between angles on the Earth's surface and on the ellipsoid. This is still of practical significance in modern geodesy, such as the reduction of the triangulateration in an engineering control network. While setting up the azimuth survey monument at the space TT&C (tracking, telemetering, and command) stations, one needs to convert the difference in geodetic azimuths

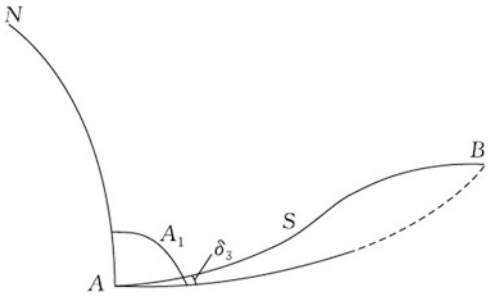


Fig. 5.33 Correction from normal section to geodesic

measured by GPS to the angles with respect to the plumb line and convert the direction of the center of the station mark to the direction of the target point.

In real operations, it is regulated that the reduction is accurate to 0.001″ in the first-order triangulation, 0.01″ in the second-order triangulation, and 0.1″ in the third- and fourth-order triangulations. Hence, not all three corrections for horizontal directions need to be computed for triangulation in all order types. They are based on the magnitude of the numerical values of each correction and the requirements for the decimal place. This is summarized in Table 5.5.

Sample computations of the three corrections for horizontal directions are as follows:

1. **Computational Formulae**

$$\left.\begin{aligned}
&\delta_1'' = -\left(\xi'' \sin A_1 - \eta'' \cos A_1\right) \tan \alpha \\
&\delta_2'' = \frac{e^2 H_2 \rho''}{2M_2} \cos^2 B_2 \sin 2A_1 \\
&\delta_3'' = -\frac{e^2 S^2 \rho''}{12N_1^2} \cos^2 B_1 \sin 2A_1 \\
&V = \sqrt{1 + e'^2 \cos^2 B} \\
&N = \frac{c}{V} \\
&M = \frac{N}{V^2} \\
&\Sigma\delta = \delta_1 + \delta_2 + \delta_3
\end{aligned}\right\} \tag{5.58}$$

2. **Example**

In Fig. 5.34 is a schematic view of the triangulation network formed by Wang Village, Zhang Village, Long Mountain, and Gao Mountain. The known data; and the values of observed directions in the computations of three corrections for horizontal directions are given in Tables 5.6 and 5.7) (Krassowski Ellipsoid, Beijing Coordinate System 1954 adopted)

Table 5.5 Computation of the three corrections for horizontal directions

Three corrections	Meaning of reduction	Primary quantities in relations	General numerical value (″)	First-order	Second–order	Third- and fourth-order
δ_1	Reduction of the observed direction value to the ellipsoidal normal	ξ, η	0.05–0.1	Applied	Applied	Handle with discretion
δ_2	Reduction of the observed direction value to the normal section direction on the ellipsoid	H_2	0.01–0.7	Applied	Applied	Handle with discretion
δ_3	Reduction of the normal section direction value to the geodesic direction	S	0.001–0.007	Applied	Not applied	Not applied

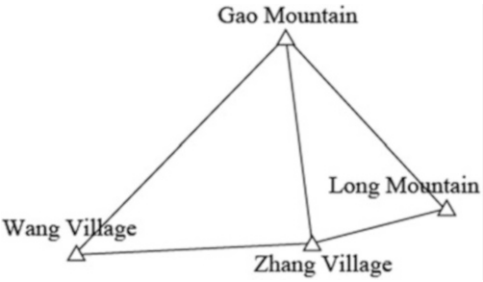


Fig. 5.34 Schematic view of triangulation network

5.5.3 *Reduction of the Observed Zenith Distance*

Zenith distance is needed for calculating the ellipsoidal height differences between two neighboring points on the Earth's surface by means of trigonometric leveling. When we observe at a triangulation point with respect to the plumb line, we get the astronomical zenith distance z_1. When the height difference is computed on the ellipsoid referring to the ellipsoidal normal, the geodetic zenith distance z is used. There are discrepancies between these two zenith distances due to the effect of the deflection of the vertical, hence reduction is needed. Let $z–z_1 = \varepsilon$, then:

$$z = z_1 + \varepsilon, \tag{5.59}$$

where ε is known as the correction for deflection of the vertical of the observed zenith distance. Below are derivations of the expression of ε.

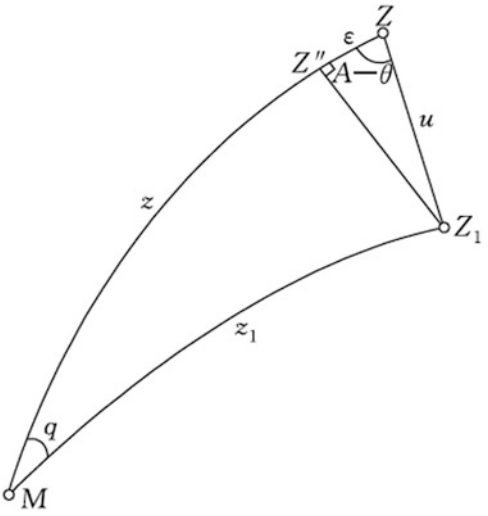
Table 5.6 Datasheet for the example triangulation network

Observation point	Normal height of the target point $H_N + a$ (m)	Height anomaly ζ (m)	Deflection of the vertical ξ (″)	Deflection of the vertical η (″)	Geodetic latitude B (º, ′)	Direction	Value of the observed direction reduced to the center of the survey mark (º, ′, ″)	Side length S (km)	Geodetic azimuth A (º, ′)	Vertical angle 90-z (′, ″)
Gao Mountain	3,579.8	+26.6	+5.5	−2.5	32, 31	Long Mountain	0, 00, 00.00	30.7	128, 17	−16, 51
						Zhang Village	40, 43, 53.34	22.0	169, 01	+23, 02
						Wang Village	102, 36, 11.45	38.1	230, 53	+23, 02
Long Mountain	3,494.9	+29.1	+7.6	−1.6	32, 20	Zhang Village	0, 00, 00.00	20.1	262, 42	+40, 54
						Gao Mountain	45, 43, 17.99	30.7	308, 25	+02, 04
Zhang Village	3,759.2	+29.3	+6.8	−1.7	32, 19	Wang Village	0, 00, 00.00	33.8	265, 58	+19, 54
						Gao Mountain	83, 04, 05.16	22.0	349, 02	−33, 00
						Long Mountain	176, 36, 54.63	20.1	82, 35	−49, 51
Wang Village	3,931.7	+29.2	+6.2	−1.9	32, 17	Gao Mountain	0, 00, 00.00	38.1	50, 26	−40, 54
						Zhang Village	35, 03, 37.45	33.8	85, 30	−25, 06

Table 5.7 Reduction of values of the observed directions to the ellipsoid

Observation point	Direction	Value of the observed direction reduced to center of the survey mark (°, ′, ″)	Three corrections for horizontal directions					Direction value on the ellipsoid L_0 (°, ′, ″)
			δ_1 (″)	δ_2 (″)	δ_3 (″)	$\Sigma\delta$ (″)	Rounded to zero (″)	
Gao Mountain	Long Mountain	0, 00, 00.00	+0.014	−0.266	+0.002	−0.250	0.000	0, 00, 00.00
	Zhang Village	40, 43, 53.34	+0.009	−0.110	0.000	−0.100	+0.150	40, 43, 53.49
	Wang Village	102, 36, 11.45	+0.039	+0.301	−0.003	+0.337	+0.587	102, 36, 12.04
Long Village	Zhang Village	0, 00, 00.00	+0.092	+0.074	0.000	+0.166	0.000	0, 00, 00.00
	Gao Mountain	45, 43, 17.99	+0.003	−0.271	+0.002	−0.267	−0.433	45, 43, 17.56
Zhang Village	Wang Village	0, 00, 00.00	+0.038	+0.043	0.000	+0.081	0.000	0, 00, 00.00
	Gao Mountain	83, 04, 05.16	+0.004	−0.104	0.000	-0.100	−0.181	83, 04, 04.98
	Long Mountain	176, 36, 54.63	+0.101	+0.070	0.000	+0.171	+0.090	176, 36, 54.72
Wang Village	Gao Mountain	0, 00, 00.00	+0.071	+0.274	−0.003	+0.342	0.000	0, 00, 00.00
	Zhang Village	35, 03, 37.45	+0.046	+0.046	0.000	+0.092	−0.250	35, 03, 37.20

Fig. 5.35 Derivations of ε



We take out the spherical triangle ZZ_1M in Fig. 5.31 (see Fig. 5.35) and draw a line Z_1Z'' perpendicular to ZM. Since q is a small quantity, $ZZ'' = \varepsilon$. Also in $\Delta ZZ_1Z''$, we see that:

$$\begin{aligned}\varepsilon &= u\cos(A-\theta)\\ &= u\cos\theta\cos A + u\sin\theta\sin A\\ &= \xi\cos A + \eta\sin A\end{aligned}$$

Hence:

$$z = z_1 + \xi\cos A + \eta\sin A. \tag{5.60}$$

This is the reduction formula of the observed zenith distance, where ξ. and η are the components of deflection of the vertical in the meridian and in the prime vertical at the observation point, respectively. A denotes the geodetic azimuth of the observed direction.

5.5.4 *Reduction of the Observed Slope Distance to the Ellipsoid*

The length observed using the rangefinder is known as the slope distance between two points on the Earth's surface. Reducing the slope distance to the ellipsoid, namely converting the slope distance D to the geodesic distance S (Fig. 5.36), is termed the reduction of the slope distance.

We will now derive the formula for the reduction of the slope distance over short distances. Two approximations are made: first, K_a and K_b are considered to coincide; and second, geodesic S is considered an arc of a great circle. Hence, the reduction in Fig. 5.36 becomes finding the solution of the plane triangle in Fig. 5.37.

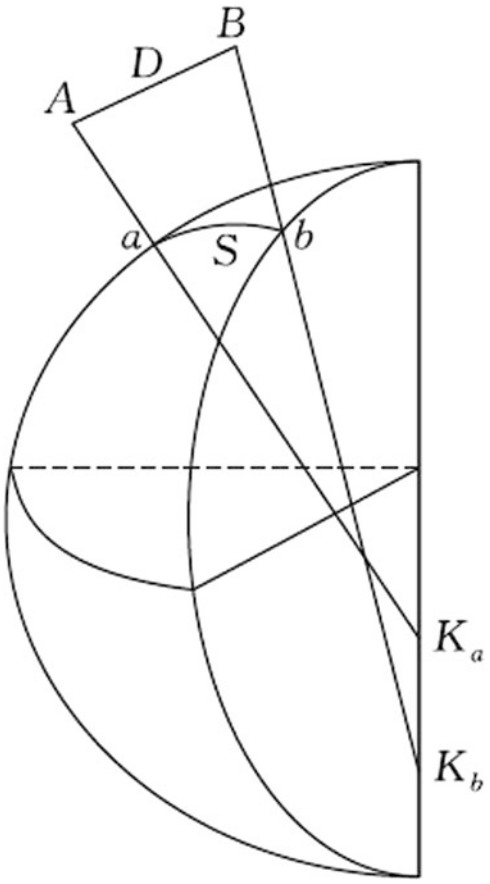


Fig. 5.36 Reduction of slope distance

Taking into consideration the error terms created by the above two approximations, we can derive the formula for reduction of the slope distance over long distances.

In Fig. 5.37:

$$S = R_A \sigma. \tag{5.61}$$

Applying the law of cosines to σ gives:

$$D^2 = (R_A + H_1)^2 + (R_A + H_2)^2 - 2(R_A + H_1)(R_A + H_2)\cos\sigma. \tag{5.62}$$

With the above two equations, the slope distance reduction can be programmed and completed. These two equations can be further rearranged.

With $\cos\sigma = 1 - 2\sin^2\frac{\sigma}{2}$, (5.62) can be rewritten as:

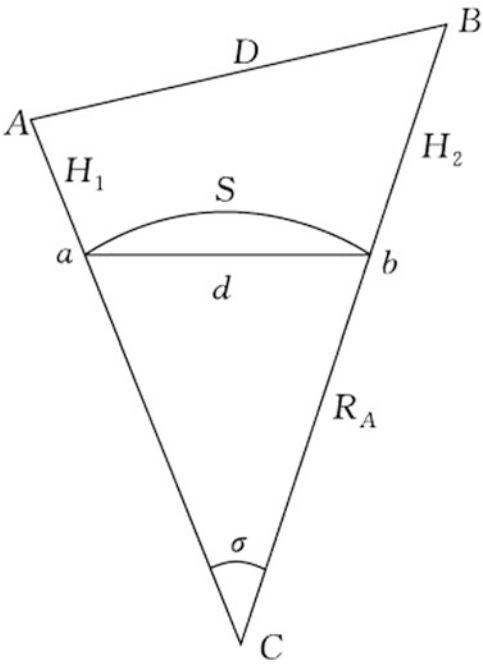


Fig. 5.37 Approximate relations for reduction of the slope distance

$$D^2 = (H_2 - H_1)^2 + 4R_A^2\left(1 + \frac{H_1}{R_A}\right)\left(1 + \frac{H_2}{R_A}\right)\sin^2\frac{\sigma}{2}.$$

Setting the chord length $d = 2R_A \sin\frac{\sigma}{2}$ and $\Delta H = H_2 - H_1$, and substituting gives:

$$D^2 = \Delta H^2 + \left(1 + \frac{H_1}{R_A}\right)\left(1 + \frac{H_2}{R_A}\right)d^2.$$

Hence, $S = R_A\sigma = 2R_A \sin^{-1}\frac{d}{2R_A}$.

Expanding the above arcsine function based on Taylor series and rearranging produces:

$$S = d + \frac{d^3}{24{R_A}^2}, \tag{5.63}$$

where

$$d = \sqrt{\frac{D^2 - \Delta H^2}{\left(1 + \frac{H_1}{R_A}\right)\left(1 + \frac{H_2}{R_A}\right)}} = R_A \cdot \sqrt{\frac{D^2 - \Delta H^2}{(R_A + H_1)(R_A + H_2)}}.$$

In practical cases, the precise reduction formula will be applied as:

Table 5.8 Computation for reduction of the slope distance (GRS80 Ellipsoid)

Elements	Example 1	Example 2
D	5,432.321 m	9,876.543 m
H_1	826.93 m	4,254.23 m
H_2	837.65 m	4,876.47 m
B	36°42′	32°12′
A	63°47′	120°24′
S	5,431.600 m	9,849.871 m

$$S = \frac{D' R_A}{R_A + H_m} + \frac{D^3}{24 {R_A}^2} + 1.25 \times 10^{-16} H_m D^2 \sin 2B \cos A, \tag{5.64}$$

where

$$D' = \sqrt{D^2 - (H_2 - H_1)^2}, H_m = \frac{1}{2}(H_1 + H_2), N = \frac{c}{\sqrt{1 + {e'}^2 \cos^2 B}},$$

$$R_A = \frac{N}{1 + {e'}^2 \cos^2 B \cos^2 A},$$

D is the known slope distance, accurate to 0.001 m; H_1 and H_2 are the ellipsoidal heights at the two end points of the observed distance, accurate to 0.001 m; B is the geodetic latitude of the origin point of the observed distance, accurate to zero decimal place; A is the geodetic azimuth of the observed distance, accurate to zero decimal place; and S is the geodesic distance on the ellipsoid that the slope distance being reduced to, accurate to 0.001 m. For an example of computation see Table 5.8.

5.5.5 *Relationship Between Astronomical Longitude and Latitude and Geodetic Longitude and Latitude (Formula for Deflection of the Vertical)*

The two components of the deflection of the vertical, ξ and η, are the two required quantities for reducing the terrestrial observation elements to the ellipsoid. According to the definition of astronomical longitude and latitude, λ and φ determine the direction of the plumb line at a given point while L and B determine the direction of the ellipsoidal normal at this point. Thus, ξ and η can be defined by the four parameters λ, φ, L, and B.

Given the fact that the geodetic latitude B is the angle between the equatorial plane and the line that is normal to the reference ellipsoid, the angle between the ellipsoidal normal and the minor axis of the ellipsoid is $90° - B$; namely, in Fig. 5.31, $PZ = 90° - B$. Similarly, $PZ_1 = 90° - \varphi$. PZ is the geodetic meridian plane and PZ_1 is the astronomical meridian plane; hence, the angle between the two meridian planes is $\lambda - L$. It is also known that $PZ' = 90° - B - \xi$. For all the quantities above, see Fig. 5.38.

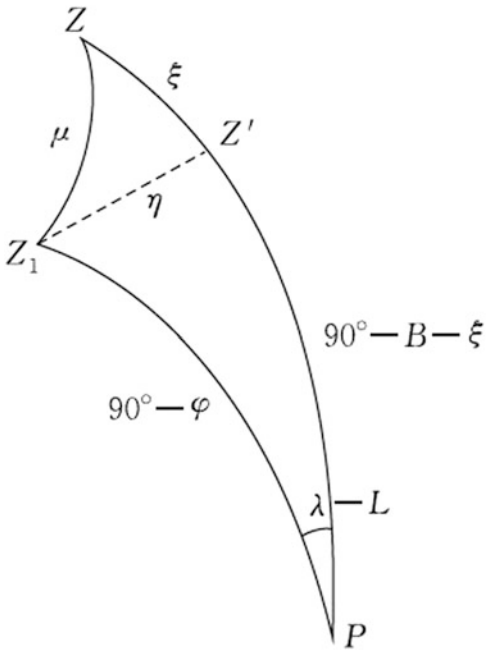


Fig. 5.38 Derivation of the formula for the deflection of the vertical

Here are two preconditions: (1) The minor axis of the ellipsoid is specified parallel to the Earth's rotational axis, translating the two axes to point A in Fig. 5.31, and they will intersect the auxiliary sphere at P; (2) Initial geodetic meridian plane and astronomical meridian plane are parallel, so the angle between the planes of the geodetic and astronomical meridians on the auxiliary sphere will be $\lambda - L$.

In Fig. 5.38, solving the spherical triangle $Z_1Z'P$ by Napier's rules gives:

$$\begin{aligned} \cos(\lambda - L) &= \tan(90^\circ - B - \xi)\tan\varphi, \\ \sin\eta &= \sin(\lambda - L)\cos\varphi, \end{aligned}$$

where both (λ–L) and η are small quantities. Putting $\sin(\lambda - L) = \lambda - L$, $\cos(\lambda - L) = 1$, $\sin\eta = \eta$, and substituting into the above equation yields:

$$\left.\begin{aligned} \xi &= \varphi - \mathrm{B} \\ \eta &= (\lambda - \mathrm{L})\cos\varphi \end{aligned}\right\}. \tag{5.65}$$

These are the formulae for the deflection of the vertical.

Given the astronomical longitude and latitude as well as geodetic longitude and latitude at a certain point, the components of the deflection of the vertical, ξ and η, at this point can be computed using the astronomical and geodetic coordinates. Therefore, it is also known as the astro-geodetic deflection of the vertical.

The astro-geodetic network formed by first-order triangulation chains (traverse) in China will determine the astronomical longitude and latitude at certain distance intervals, intended to compute deflection of the vertical to serve the needs of

reducing the observed direction, as well as other applications . The deflection of the vertical can also be obtained using the gravimetric data, known as the gravimetric deflection of the vertical (also referred to as absolute deflection of the vertical). It is defined relative to the normal ellipsoid. The astro-geodetic deflection of the vertical (also known as relative deflection of the vertical) is defined relative to the reference ellipsoid. As with the normal ellipsoid, the reference ellipsoid in modern geodesy adopts the geocentric orientation. Hence, the distinctions between the absolute deflection of the vertical and the relative deflection of the vertical no longer exist.

It is known that the reduction of the observed directions requires the deflection of the vertical at every geodetic point. Nevertheless, in practical cases it is impossible to carry out an astronomical survey at every single geodetic point. Hence, we may take the following measures to solve the problem. When there is a lack of gravimetric data, a linear interpolation approach can be employed based on the known values of deflections of the vertical at some point. Of course, this is not very realistic in reality. In mountainous areas, deflections of the vertical will have large nonlinear changes. Such nonlinear deflections will exist even in plane areas due to the non-homogeneous distribution of mass inside the Earth. Hence, the data obtained from linear interpolation are not accurate and even display quite large discrepancies. Gravity data are needed to achieve the required accuracy of the astro-geodetic deflection of the vertical. They can be utilized to obtain gravimetric deflection of the vertical that can be converted to the astro-geodetic deflection of the vertical. Therefore, determining the deflection of the vertical at a given surface point involves combinations of the astronomical, geodetic, and gravimetric data.

Analogous to the grid model of height anomaly (Sect. 4.4), we can establish the grid of deflection of the vertical in a certain area, including the grid model of the component of deflections of the vertical in the meridian and in the prime vertical, respectively. A grid model of the deflection of the vertical is the discretized numerical representation of the vertical deflection components within a certain range, and is an aggregation of the values of vertical deflection components of all the uniformly spaced grid points within this range, stored in the database in the form of grid data structure. In mobile missile warfare, one needs to know the vertical deflection at any point within the mobile area. As a result, before war, the grid model of deflections of the vertical in the mobile area needs to be established.

It follows from (5.65) that:

$$\left.\begin{aligned} B &= \varphi - \xi \\ L &= \lambda - \eta \sec \varphi \end{aligned}\right\}. \tag{5.66}$$

The above is the relational expression between astronomical longitude and latitude and geodetic longitude and latitude. From the expression, theoretically, given the astronomical longitude and latitude of a point and deflection of the vertical of this point, one can obtain the geodetic longitude and latitude of the point, yet the accuracy of L and B obtained by this method is quite low and is therefore not used in practical cases. We can make the following analyses. The mean square observational errors of astronomical longitude and latitude are $m_\lambda =$

$\pm$ 0.02s $= \pm$ 0. 3″ and $m_\varphi = \pm$ 0. 3″, corresponding to a terrestrial distance of $\pm$9 m, and the mean square error of the point is $\pm 9\text{m} \times \sqrt{2} = \pm 12.6\text{m}$. Correcting the deflection of the vertical to within $\pm 1''$ with gravimetric data, the mean square error of the point becomes $\pm 1'' \times \sqrt{2} = \pm 1.4''$, corresponding to a distance of $\pm$42 m. Taking the two aspects into consideration, the effect is $\pm$44 m.

5.5.6 *Relationship Between Astronomical Azimuth and Geodetic Azimuth (Laplace Azimuth Formula)*

One of the applications of an astronomical survey is to determine geodetic azimuths in order to define the orientation of the control network and control the accumulation of azimuth errors. Hence, the observed astronomical azimuth should be reduced to the geodetic azimuth.

From Fig. 5.31, the astronomical azimuth of the *AM* direction is:

$$\alpha = \theta_1 + R_1.$$

The geodetic azimuth of the *AM* direction is:

$$A = \theta + R.$$

The two equations subtracted from each other give:

$$\alpha - A = (\theta_1 - \theta) + (R_1 - R),$$

where $R_1 - R$ is the correction for deflection of the vertical of the observed direction, namely (5.55)

To obtain the expression of $\theta_1 - \theta$, in $P\theta\theta_1$ in Fig. 5.31, applying Napier's rules in the right-angled spherical triangle gives:

$$\sin\varphi = \tan(\theta_1 - \theta)\tan(90^\circ - (\lambda - L)) = (\theta_1 - \theta)\cot(\lambda - L)$$

Hence, we get:

$$\theta_1 - \theta = (\lambda - L)\sin\varphi,$$

and

$$A = \alpha - (\lambda - L)\sin\varphi - (\xi \sin A - \eta \cos A)\cot z_1.$$

The last term at the right-hand side of the above equation is known as the term of correction for the deflection of the vertical. In general, its value is only a few hundredths of a second or even less. In the first-order astronomical survey, the mean

square observational error of astronomical azimuth is ± 0.5 ’ ’, so the effect of the vertical deflection is far less than the observational error of astronomical azimuth, which can be neglected. Inserting:

$$A = \alpha - (\lambda - L)\sin\varphi, \tag{5.67}$$

into (5.65), the above equation can be written as:

$$A = \alpha - \eta\tan\varphi. \tag{5.68}$$

Equations (5.67) and (5.68) are formulae for the reduction of the astronomical azimuth, known as the Laplace azimuth formula. The geodetic azimuth reduced by this formula is known as the Laplace azimuth, also referred to as the initial geodetic azimuth.

Applying the law of error propagation to (5.67) yields:

$$m_A{}^2 = m_\alpha{}^2 + \sin^2\varphi \cdot m_\lambda{}^2 + \sin^2\varphi \cdot m_L{}^2 + \left(\frac{\lambda - L}{\rho}\right)^2 \cdot \cos^2\varphi \cdot m_\varphi{}^2.$$

The last two terms in the above equation are small quantities, and neglecting them gives:

$$m_A{}^2 = m_\alpha{}^2 + \sin^2\varphi \cdot m_\lambda{}^2.$$

With $\varphi = 30^\circ$, and substituting $m_\alpha = \pm\ 0.\ 5''$ and $m_\lambda = \pm\ 0.\ 3''$ into the above equation, yields $m_A = \pm\ 0.\ 6''$. In this case, the accuracy of Laplace azimuth is approximately $\pm\ 0.\ 6''$.

In the geodetic control network, the geodetic azimuth of each point is obtained through pointwise calculations, which can be affected by the accumulation of angle observation errors. For instance, the single chain of triangles provides one route of 16 sides through which computations are carried out. The orientation error of each side is $\pm\ 0.\ 5''$. Then, the azimuth error of the last side is $\pm 0.5'' \sqrt{16} = \pm 2.0''$. The mean square error of the Laplace azimuth is approximately $\pm\ 0.\ 6''$. Apparently, the method can achieve a higher accuracy than the pointwise calculation.

Hence, in the classical geodetic control network, an astronomical survey is carried out at certain distance intervals to compute the Laplace azimuth. This can help to control the accumulation of azimuth errors in the geodetic network.

5.6 Relationship Between the Geodetic Coordinate System and the Geodesic Polar Coordinate System

5.6.1 Geodesic Polar Coordinate Systems and the Solution of Geodetic Problems

A geodesic polar coordinate system is established on the surface of the ellipsoid. The position of a point on the ellipsoid is represented by the geodesic distance S and geodetic azimuth A from the polar point to this point of interest, as shown in Fig. 5.39. Let P_1 be the polar point on the ellipsoid, the meridian P_1N that passes through P_1 be the polar axis, the geodesic distance S that connects P_1 and P (to be computed) be the polar radius, and the geodetic azimuth A of the geodesic at point P_1 be the polar angle; then the location of point P on the ellipsoid is expressed by (S, A).

The geodesic polar coordinate system is used to show the relative horizontal positions between two points on the ellipsoid, often applied in the case where solution of relative positions is needed for long-distance weapon launching or navigation.

Calculations of geodetic coordinates of an unknown point on the ellipsoid based on the observed angles and distances using geodetic surveying or calculations of the geodesic distance and geodetic azimuth between two points based on their geodetic coordinates are known as solutions of geodetic problems, calculations of geodetic coordinates, or calculations of geodetic positions. The solutions of geodetic problems include both direct and inverse solutions.

In Fig. 5.40, given the geodetic coordinates (L_1,B_1) of point P_1, the geodesic distance S from point P_1 to P_2, and the geodetic azimuth A_1 from P_1 to P_2, the direct solution of the geodetic problem provides the geodetic coordinates (L_2,B_2) of point P_2 and the reverse azimuth A_2 of the geodesic at point P_2. Given the geodetic coordinates (L_1,B_1) and (L_2,B_2) of P_1 and P_2, the inverse solution of the geodetic problem is required to find the forward and reverse azimuths A_1, A_2 and the geodesic distance S of P_1 and P_2. From the definition of the geodesic polar coordinates, (S, A_1) and (S, A_2) are the geodesic polar coordinates of points P_2 and P_1, respectively. Hence, the solution of geodetic problems is the interconversion between the geodetic coordinates and geodesic polar coordinates.

The solution of geodetic problems can be applied in many ways. It can be used to calculate the geodetic coordinates on the ellipsoid in geodetic surveying (as shown in Fig. 5.29). Apart from that, with the advancement of modern spatial technology, aviation, and navigation, the solution of geodetic problems (particularly the inverse solution of geodetic problems) is playing a more prominent role. These different applications and requirements also generate different methods and formulae for the solution of geodetic problems.

Computing geodetic coordinates on the ellipsoid is far more complicated than computing coordinates on the plane due to the significantly more complex

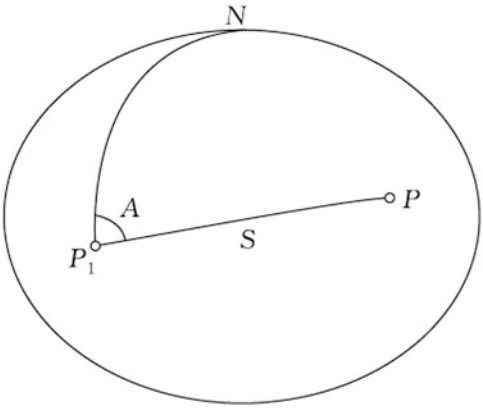


Fig. 5.39 Geodesic polar coordinate system

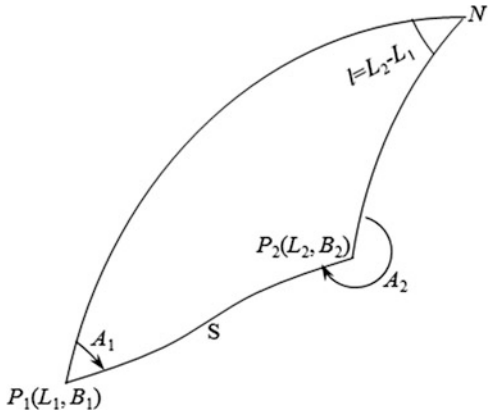


Fig. 5.40 Solution of geodetic problems

mathematical properties of the ellipsoid than the plane. Because of this complexity, there are various formulae, dozens at present, for solving geodetic problems. Based on distance, the formulae can generally be categorized into short-distance (within 400 km), mid-distance (400–1,600 km), and long-distance (1,000–20,000 km). Based on accuracy, they can be sorted into precise formulae and approximate formulae.

Theoretically, the formulae for solutions of geodetic problems are mostly based on the three differential equations of geodesics, although their forms and methods of derivations vary. For solution of long-distance geodetic problems, Clairaut's equation for geodesics should also be applied. The formulae for solutions can basically be classified into the following three categories in light of the methods for solution.

1. Based on the geodesic on the ellipsoid and its three differential equations, we expand the differences in geodetic longitude l, geodetic latitude b, and geodetic azimuth a of the two endpoints of the geodesic into the ascending power series of

the geodesic distance S. The features of this type of formulae are that the accuracy of solution is distance-dependent, and the longer the distance the slower the rates of convergence. Converge may not even occur, so finding solutions would be impossible. Hence, this approach is better applicable to short distances.

2. We take an auxiliary sphere, convert the ellipsoidal elements to the spherical surface, solve the problem by applying the formula for spherical triangles on a spherical surface, and then reduce the computed results back to the ellipsoid. Because the difference between the ellipsoid and spheroid is only a very small flattening, the expressions for transformation of the corresponding elements between the ellipsoidal surface and the spherical surface only involve the ascending power series of the small quantities e^2 (eccentricity squared) or e'^2. The Bessel's formula studied in this section can represent this type of formula. Such formulae are independent of distance and can be applied to the solution of geodetic problems over any distances.
3. Take advantage of numerical integration and solve the direct and inverse geodetic problems on the ellipsoid. In general, the solutions are composed of a strict solution for the sphere plus a correction to the ellipsoid, determined by numerical integration. By employing numerical integration, routines that are usually available in current computer software like MATLAB, the problems of classical geodesy are easily solved to the desired accuracy (Sjöberg et al. 2012).

The accuracy of solutions of geodetic problems depends on different practical matters. Take the adjustment of the astro-geodetic network for instance. The relative mean square error of the side lengths in the first-order triangulation chain $\frac{m_S}{S} = 1 : 200000$ and the accuracy of azimuth $m_A = \pm\ 1.\ 0''$, i.e., $\frac{m_A}{\rho} \approx 1 : 200000$. We set $S = 20$ km and let the coordinate components be the same as the longitudinal and lateral errors, namely:

$$m_x = m_y = \pm 0.1 \text{ m}.$$

After considering the adjustment of the first-order triangulation chain, the accuracy of the point positioning can be slightly improved. Putting $m_x = m_y = 0.09$ m, if expressed by geodetic coordinates, when $B = 45°$, we obtain:

$$m_B = \frac{m_x}{M}\rho'' = \pm 0.003'',$$
$$m_L = \frac{m_y}{N\cos B}\rho'' = \pm 0.004''.$$

When deciding accuracy requirements for the solution formulae, one should generally adhere to the following principle: it has to be ensured that the computational errors introduced by the formulae do not have any further effect on the real accuracy of the field observations and adjusted values. In accordance with this principle, considering the possible accumulation of errors in the pointwise

calculation of geodetic coordinates along the first-order triangulation chain, the geodetic longitude and latitude should be accurate to 0.0001″.

In the first-order triangulation, the final result of the azimuth is accurate to 0.01″. Hence, in solutions of geodetic problems the geodetic azimuth is accurate to 0.001″.

The above discussions of computational accuracy are mainly concerned with the adjustment of astro-geodetic networks and the calculation of coordinates of the first-order geodetic points. When applied in other situations, the accuracy of computation should be determined by the applications. For instance, in navigation application, the geodetic longitude, latitude, and azimuth can be accurate to 0.1″, and the distance needs to be accurate only to 10 m.

To understand the rationale for solving geodetic problems, we must first provide the method for point-by-point integration for the direct solution of geodetic problems.

We divide the length of P_1P_2 (the interval between P_1 and P_2) in Fig. 5.40 into n sections. The differences in longitude, latitude, and azimuth between the two endpoints of each small section are dL, dB, and dA, respectively, satisfying the conditions of differential equations of geodesics. Hence, the direct solution is given by:

$$\left.\begin{aligned} L_2 - L_1 &= \int_{L_1}^{L_2} \mathrm{d}L = \int_0^S \frac{\sin A}{N} \sec B \mathrm{d}S \approx \sum_{i=1}^{n} \frac{\sin A_i}{N_i} \sec B_i \Delta S_i \\ B_2 - B_1 &= \int_{B_1}^{B_2} \mathrm{d}B = \int_0^S \frac{\cos A}{M} \mathrm{d}S \approx \sum_{i=1}^{n} \frac{\cos A_i}{M_i} \Delta S_i \\ A_2 - A_1 \pm 180 &= \int_{A_1}^{A_2 \pm 180} \mathrm{d}A = \int_0^S \frac{\sin A}{N} \tan B \mathrm{d}S \approx \sum_{i=1}^{n} \frac{\sin A_i}{N_i} \tan B_i \Delta S_i \end{aligned}\right\}. \tag{5.69}$$

Repeated computations of the above equations will solve the short-distance direct geodetic problems and also enable a high degree of accuracy. The above equations have shown that as n increases, so does the accuracy of the computations .

5.6.2 *Series Expansions of the Solution of the Geodetic Problem*

In Fig. 5.40, at the given point P_1 (L_1,B_1), when the geodesic azimuth A_1 is determined, the geodetic longitude, latitude, and azimuth of an arbitrary point on the geodesic are the functions of the geodesic distance S, namely:

$$L = L(S), B = B(S), A = A(S).$$

Obviously, the above function can be differentiated repeatedly. Expanding the above function at point P_1 based on Maclaurin series results in:

$$\begin{aligned} l &= L_2 - L_1 = \left(\frac{\mathrm{d}L}{\mathrm{d}S}\right)_1 S + \left(\frac{\mathrm{d}^2 L}{\mathrm{d}S^2}\right)_1 \frac{S^2}{2} + \left(\frac{\mathrm{d}^3 L}{\mathrm{d}S^3}\right)_1 \frac{S^3}{6} + \cdots, \\ b &= B_2 - B_1 = \left(\frac{\mathrm{d}B}{\mathrm{d}S}\right)_1 S + \left(\frac{\mathrm{d}^2 B}{\mathrm{d}S^2}\right)_1 \frac{S^2}{2} + \left(\frac{\mathrm{d}^3 B}{\mathrm{d}S^3}\right)_1 \frac{S^3}{6} + \cdots, \\ a &= A_2 \mp 180^\circ - A_1 = \left(\frac{\mathrm{d}A}{\mathrm{d}S}\right)_1 S + \left(\frac{\mathrm{d}^2 A}{\mathrm{d}S^2}\right)_1 \frac{S^2}{2} + \left(\frac{\mathrm{d}^3 A}{\mathrm{d}S^3}\right)_1 \frac{S^3}{6} + \cdots. \end{aligned} \tag{5.70}$$

where the subscript "1" denotes the value when the derivatives of various orders take $S = 0$ (i.e., the value at point P_1; $B = B_1$, $A = A_1$). It can thus be seen that finding the derivatives of different orders in the equations is the key to obtain the expansions of l, b, and a. Here, three first-order derivatives will form a differential equation of the geodesic. With $N = \frac{c}{V}, M = \frac{c}{V^3}$, one obtains:

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}S} &= \frac{V}{c} \sec B \sin A, \\ \frac{\mathrm{d}B}{\mathrm{d}S} &= \frac{V^3}{c} \cos A, \\ \frac{\mathrm{d}A}{\mathrm{d}S} &= \frac{V}{c} \tan B \sin A. \end{aligned}$$

Taking repeated derivatives of the above equations results in the derivatives of various orders in (5.70); hence, one will get the power series in geodesic distance S expanded from the differences in longitude l, latitude b, and azimuth a, generally known as the Legendre series.

Legendre series converge more slowly. However, if l, b, and a are expanded into the power series in S at the midpoint of the geodesic $P_{S/2}$ instead of the end point of the geodesic P_1, then the convergence rate of the series will increase considerably. Expanding the difference in geodetic longitude at midpoint $P_{S/2}$ based on the Taylor series yields:

$$L_2 - L_{S/2} = \left(\frac{\mathrm{d}L}{\mathrm{d}S}\right)_{S/2}\left(\frac{S}{2}\right) + \frac{1}{2}\left(\frac{\mathrm{d}^2L}{\mathrm{d}S^2}\right)_{S/2}\left(\frac{S}{2}\right)^2 + \frac{1}{6}\left(\frac{\mathrm{d}^3L}{\mathrm{d}S^3}\right)_{S/2}\left(\frac{S}{2}\right)^3 + \frac{1}{24}\left(\frac{\mathrm{d}^4L}{\mathrm{d}S^4}\right)_{S/2}\left(\frac{S}{2}\right)^4 + \cdots,$$

$$L_1 - L_{S/2} = \left(\frac{\mathrm{d}L}{\mathrm{d}S}\right)_{S/2}\left(-\frac{S}{2}\right) + \frac{1}{2}\left(\frac{\mathrm{d}^2L}{\mathrm{d}S^2}\right)_{S/2}\left(-\frac{S}{2}\right)^2 + \frac{1}{6}\left(\frac{\mathrm{d}^3L}{\mathrm{d}S^3}\right)_{S/2}\left(-\frac{S}{2}\right)^3 + \frac{1}{24}\left(\frac{\mathrm{d}^4L}{\mathrm{d}S^4}\right)_{S/2}\left(-\frac{S}{2}\right)^4 + \cdots.$$

Subtracting the above two equations from each other gives:

$$l = L_2 - L_1 = \left(\frac{\mathrm{d}L}{\mathrm{d}S}\right)_{S/2} S + \frac{1}{24}\left(\frac{\mathrm{d}^3L}{\mathrm{d}S^3}\right)_{S/2} S^3 + \cdots.$$

In like manner, we can obtain the formulae for differences in geodetic latitude and azimuth. Combining them with the above expression yields:

$$\left.\begin{aligned} l = L_2 - L_1 &= \left(\frac{\mathrm{d}L}{\mathrm{d}S}\right)_{\frac{S}{2}} S + \frac{1}{24}\left(\frac{\mathrm{d}^3L}{\mathrm{d}S^3}\right)_{\frac{S}{2}} S^3 + \cdots \\ b = B_2 - B_1 &= \left(\frac{\mathrm{d}B}{\mathrm{d}S}\right)_{\frac{S}{2}} S + \frac{1}{24}\left(\frac{\mathrm{d}^3B}{\mathrm{d}S^3}\right)_{\frac{S}{2}} S^3 + \cdots \\ a = A_2 - A_1 \mp 180^{\mathrm{o}} &= \left(\frac{\mathrm{d}A}{\mathrm{d}S}\right)_{\frac{S}{2}} S + \frac{1}{24}\left(\frac{\mathrm{d}^3A}{\mathrm{d}S^3}\right)_{\frac{S}{2}} S^3 + \cdots \end{aligned}\right\}. \tag{5.71}$$

where the subscript $S/2$ indicates that the derivatives of various orders in the bracket will be taken according to the geodetic latitude $B_{S/2}$ and geodetic azimuth $A_{S/2}$ at the midpoint of the geodesic $\mathrm{P}_{S/2}$. In (5.71), although only two terms are listed, the accuracy actually reaches S^4 terms. Thus, it converges more rapidly than the Legendre series. In (5.71), the geodetic latitude $B_{S/2}$ and geodetic azimuth $A_{S/2}$ are actually unknown, so the equation cannot be solved directly, and we need to convert the derivatives of the equation. Assume that:

$$\left.\begin{aligned} B_m &= \frac{1}{2}(B_1 + B_2) \\ A_m &= \frac{1}{2}(A_1 + A_2 \mp 180^{\circ}) \end{aligned}\right\}.$$

Apparently, $B_m \neq B_{S/2}$, $A_m \neq A_{S/2}$, for the very small flattening of the ellipsoid, although the difference between them is not large. We can then derive the estimating equation of $B_m - B_{S/2}$, $A_m - A_{S/2}$ so as to convert the derivatives based on $B_{S/2}$, $A_{S/2}$ to the derivatives based on B_m, A_m, derivations omitted. B_2 and A_2 are the unknowns in the solution of direct geodetic problems, so the exact values of B_{m} and A_{m} are unknown and need to be obtained using a successive approximation.

Equation (5.71) is the formula for a direct solution of the geodetic problem, based on which one can derive the corresponding formula for an inverse solution of the geodetic problem. These formulae are improvements of the Legendre series and are applicable to the solution of short-distance geodetic problems, known as the Gauss mid-latitude formula (see, e.g., Krakiwsky and Thomson 1974).

5.6.3 *Bessel's Formula for the Solution of the Geodetic Problem*

Overview

The series expansion of the solution of the geodetic problem is to express the differences in geodetic longitude, latitude, and azimuth as a function of the geodesic distance S. It is evident that to achieve the desired accuracy, the longer the distance, the more complex the formula structure becomes and may even become unsolvable. Hence, such a formula is not suitable for solving long-distance geodetic problems.

From the spherical trigonometry, we are clear that the formulae for spherical triangles are all expressed by the trigonometric function of angles, where the accuracy of solving spherical triangles is independent of the spherical distance. In addition, the flattening of the Earth ellipsoid is very small, and when the ellipsoidal elements (longitude, latitude, side length, and azimuth) are converted into the corresponding elements on the spherical surface and represented by angles, the corresponding corrections expressed by angles will be quite small and independent of distances. Hence, the general approach to solving the long-distance geodetic problem is to use a spherical surface as a bridge, which means to establish relations of projection between the ellipsoidal elements and the corresponding spherical elements under certain projection conditions, to establish the precise relations between the elements on the spherical surface using the formula for spherical

triangles, and eventually to transform the spherical elements computed back into the ellipsoidal elements.

Clearly, the key issue is to establish the corresponding relationship between the ellipsoidal elements and the spherical elements. From Fig. 5.40, the corresponding ellipsoidal and spherical elements should include the six elements B_1, B_2, A_1, A_2, l, S.

Reduced Latitude

When the elements on the ellipsoidal surface are converted into the corresponding elements on the spherical surface, we may obtain different formulae for solutions of the geodetic problem over long distances due to different options for the projection conditions or different methods of integration. This section will discuss the representative formula, i.e., the formula for the solution of the geodetic problem over long distances put forward by Bessel in 1825 (see, e.g., Krakiwsky and Thomson 1974). Our discussion first considers reduced latitude and the transformation between the reduced latitude and the geodetic latitude.

As shown in Fig. 5.41, *NPSN* represents a meridian ellipse. Create an auxiliary circle with its center at the center of the ellipse O, and the equatorial radius a as its radius. Extend the ordinate line $P'P$ of point P to intersect the circle at P''. Join $P''O$; then $\angle\ P''OP'$ is known as the reduced latitude of point P, denoted by u.

Apparently, for any point P on the meridian ellipse there will be a reduced latitude u corresponding to it. The relations of transformation between the reduced latitude and geodetic latitude are derived as follows:

We establish a right-angled plane coordinate system XOY in Fig. 5.41, and the coordinates of P are given by:

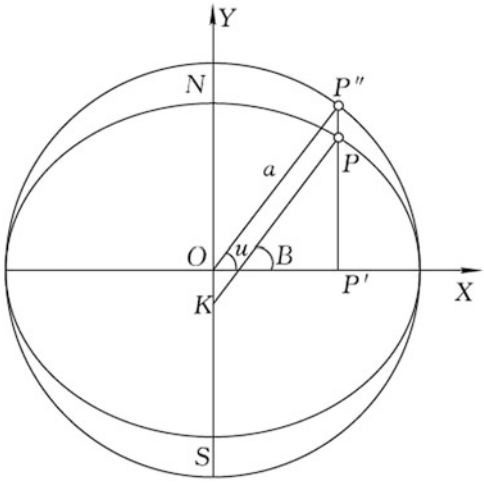
$$\left.\begin{aligned} X &= \frac{a}{W}\cos B \\ Y &= \frac{a}{W}\left(1-e^2\right)\sin B = \frac{b}{W}\sqrt{1-e^2}\sin B \end{aligned}\right\}. \tag{5.72}$$

It is known from Fig. 5.41 that $X = a\cos u$, substituting X into the elliptical equation $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ results in Y, hence:

$$\left.\begin{aligned} X &= a\cos u \\ Y &= b\sin u \end{aligned}\right\}. \tag{5.73}$$

Comparing (5.72) and (5.73) yields:

Fig. 5.41 Reduced latitude u



$$\left.\begin{aligned} \cos u &= \frac{1}{W}\cos B \\ \sin u &= \frac{\sqrt{1-e^2}}{W}\sin B \end{aligned}\right\}. \tag{5.74}$$

Introducing $W = V\sqrt{1-e^2}$, we get:

$$\left.\begin{aligned} \cos u &= \frac{1}{V\sqrt{1-e^2}}\cos B \\ \sin u &= \frac{1}{V}\sin B \end{aligned}\right\}. \tag{5.75}$$

Dividing the two equations in (5.74) or (5.75) by each other results in:

$$\tan u = \sqrt{1-e^2}\tan B. \tag{5.76}$$

Equation (5.76) is the formula for transformation between the reduced latitude u and geodetic latitude B. It follows from (5.76) that the reduced latitude u and geodetic latitude B of the same point have a specified corresponding relationship. Generally, geodetic latitude is greater than the reduced latitude.

Taking the derivative of (5.76) results in the differential relations between u and B as follows:

$$\frac{\mathrm{d}u}{\cos^2 u} = \frac{\sqrt{1-e^2}}{\cos^2 B}\mathrm{d}B,$$

With (5.75), one obtains:

$$\frac{\mathrm{d}B}{\mathrm{d}u} = V^2\sqrt{1-e^2}. \tag{5.77}$$

Underlying Principle of Bessel's Solution of the Geodetic Problem

Bessel's formula for the solution of the geodetic problem is first to create an auxiliary sphere with its center at the center of the ellipsoid and the radius of any length (the problem of spherical triangles is independent of the length of radius) and then to solve according to the three steps below:

1. Project the ellipsoidal elements onto the spherical surface according to certain conditions
2. Solve the geodetic problem on the spherical surface
3. Convert the spherical elements obtained into the corresponding ellipsoidal elements based on the relations of projection

The three conditions of projection for Bessel's solution of the geodetic problem are as follows:

1. The spherical latitude of the point on the spheroid after projection is equal to the reduced latitude of the corresponding point on the ellipsoid
2. Projection of the geodesic between two points on the ellipsoid onto the auxiliary sphere is a great circle
3. The numerical value of the geodetic azimuth A_1 remains unchanged after projection

In Fig. 5.42, after the auxiliary sphere is created and the projection is done according to the three conditions above, there is a geodetic polar triangle NP_1P_2 on the ellipsoid, and a specified polar triangle $N'P_1'P_2'$ will, correspondingly, be on the spherical surface, where $N'P_1' = 90^\circ - u_1$, $N'P_2' = 90^\circ - u_2$, σ is the great circle, and $\angle N'P_1'P_2' = A_1$. Let the forward azimuth of the geodesic P_1P_2 at the point P_2 be A_2' and the forward azimuth of the great circle $P_1'P_2'$ at the point P_2' be α_2'. Applying the sine theorem to the spherical triangle $N'P_1'P_2'$ results in:

$$\cos u_1 \sin A_1 = \cos u_2 \sin \alpha_2'. \tag{5.78}$$

According to Clairaut's equation for geodesics, $r \sin A = C$, and with (5.73) $r = x = a \cos u$, one obtains:

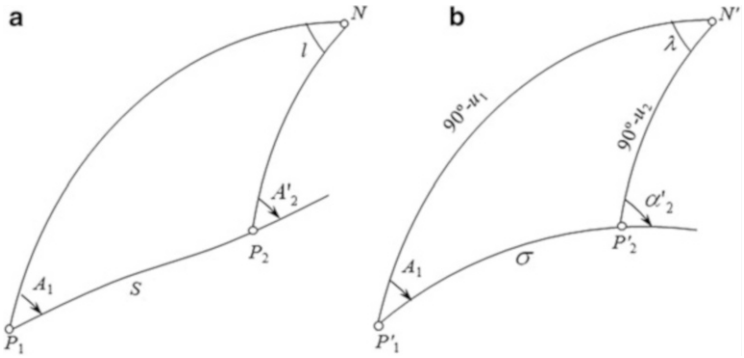


Fig. 5.42 Relationship between ellipsoidal (**a**) and spherical (**b**) projections

$$\cos u \sin A_1 = C.$$

Then we see that:

$$\cos u_1 \sin A_1 = \cos u_2 \sin A_2^{'}. \tag{5.79}$$

Comparing (5.78) and (5.79), we have:

$$\alpha_2^{'} = A_2^{'}. \tag{5.80}$$

The above equation shows that, in Bessel's solution of geodetic problems, the azimuth of this geodesic remains the same after projection.

Four elements (u_1, u_2, A_1, and A_2) of the six corresponding elements on the ellipsoid and the spherical surface have so far been determined, and the rest are not yet known, including the relationship between λ and l, and between σ and S. Hence, the expressions for the meridian arc elements and the parallel arc elements on the ellipsoid and the auxiliary sphere according to the differential equations of geodesics can be written on the ellipsoid as:

$$\left.\begin{aligned} \mathrm{d}S\cos A &= M\mathrm{d}B \\ \mathrm{d}S\sin A &= N\cos B\mathrm{d}l \end{aligned}\right\}, \tag{5.81}$$

and on the auxiliary sphere as:

$$\left.\begin{aligned} \mathrm{d}\sigma\cos A &= \mathrm{d}u \\ \mathrm{d}\sigma\sin A &= \cos u\mathrm{d}\lambda \end{aligned}\right\}, \tag{5.82}$$

where $\mathrm{d}\sigma$ is measured by angle unit.

It follows from the above two sets of formulae that:

$$\frac{\mathrm{d}S}{\mathrm{d}\sigma} = M\frac{\mathrm{d}B}{\mathrm{d}u},$$
$$\frac{\mathrm{d}S}{\mathrm{d}\sigma} = N\frac{\cos B}{\cos u}\frac{\mathrm{d}l}{\mathrm{d}\lambda}.$$

With $M = \dfrac{a(1-e^2)}{W^3} = \dfrac{a}{V^3\sqrt{1-e^2}}$, $N = \frac{a}{W}$, and (5.77) and (5.74), one obtains:

$$\mathrm{d}S = \frac{a}{V}\mathrm{d}\sigma, \tag{5.83}$$

$$\mathrm{d}l = \frac{1}{V}\mathrm{d}\lambda. \tag{5.84}$$

With $\cos u = \dfrac{1}{V\sqrt{1-e^2}}\cos B$,
and from:

$$V^2 = 1 + e'^2\cos^2 B = 1 + e'^2\cos^2 uV^2\left(1-e^2\right) = 1 + e^2V^2\cos^2 u,$$

with

$$V = \frac{1}{\sqrt{1-e^2\cos^2 u}}$$

substituted into (5.83)and (5.84) we obtain:

$$\mathrm{d}S = a\sqrt{1-e^2\cos^2 u}\mathrm{d}\sigma, \tag{5.85}$$

$$\mathrm{d}l = \sqrt{1-e^2\cos^2 u}\mathrm{d}\lambda. \tag{5.86}$$

Equations (5.85) and (5.86) are the differential equations that define the relationship between the side length and longitude difference on the ellipsoid and the corresponding side length and longitude difference on the auxiliary sphere. They are known as the Bessel's differential equation. The relations between S and σ and between l and λ can be obtained by solving this set of differential equations. Applying different integration methods will result in different formulae, which distinguishes Bessel's formula for the solution of the geodetic problem from many other formulae for solving the problem over long distances.

Bessel's formula applies to any distances because its integration is not in the power series of the side length, longitude, or latitude (differences of longitude or latitude), but in the power series of the eccentricity squared e^2 (or e'^2).

Solution of Bessel's Differential Equation

The Relationship Between S and σ

Integrating (5.85) gives the relationship between S and σ. In order to find the integral, we shall first convert u into a function of σ as shown in Fig. 5.43. We extend the arc of the great circle $P_1'P_2'$ to intersect the equator of the auxiliary sphere at point P_0'. The azimuth of $P_1'P_2'$ at the point P_0' is m, and the arc of the great circle $P_0'P_1' = M$. Obviously, when point $P_1'$and the arc of the great circle $P_1'P_2'$ are given, the values of m and M are also defined. So, the purpose of extending the arc of the great circle $P_1'P_2'$ is to find the spherical triangle that is relevant to the spherical quadrangle in order to apply the formula for spherical triangles to find the solution.

We assume that P' is a moving point along the arc $P_1'P_2'$. When P' moves, the distance σ between P' and P_1' and the spherical latitude u of point P' also change accordingly; hence, the relationship between u and σ can be established.

Considering P_2' as the moving point P', from the right-angled spherical triangle $P_0'Q_2P_2'$, one obtains:

$$\sin u = \cos m \ \sin(M+\sigma)$$

or

$$\cos^2 u = 1 - \cos^2 m \ \sin^2(M+\sigma). \tag{5.87}$$

Substituting the above equation into (5.85) gives:

$$\begin{aligned} \mathrm{d}S &= a\sqrt{1-e^2+e^2\cos^2 m\sin^2(M+\sigma)}\mathrm{d}\sigma \\ &= a\sqrt{1-e^2}\sqrt{1+\frac{e^2}{1-e^2}\cos^2 m\sin^2(M+\sigma)}\mathrm{d}\sigma \\ &= b\sqrt{1+e'^2\cos^2 m\sin^2(M+\sigma)}\mathrm{d}\sigma, \end{aligned}$$

which can be written as:

$$\mathrm{d}S = b\sqrt{1+k^2\sin^2(M+\sigma)}\mathrm{d}\sigma, \tag{5.88}$$

where $k^2 = e'^2\cos^2 m$.

In order to find the integral of (5.88), we expand the integrand into a series of σ as follows:

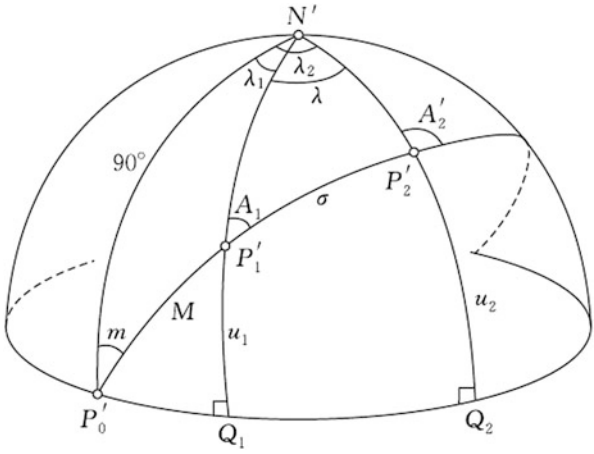


Fig. 5.43 The relationship between u and σ

$$\sqrt{1+k^2\sin^2(M+\sigma)} = 1+\frac{1}{2}k^2\sin^2(M+\sigma) - \frac{1}{8}k^4\sin^4(M+\sigma)+\cdots,$$

With

$$\sin^2(M+\sigma) = \frac{1}{2} - \frac{1}{2}\cos 2(M+\sigma),$$
$$\sin^4(M+\sigma) = \frac{3}{8} - \frac{1}{2}\cos 2\big(M+\sigma\big) + \frac{1}{8}\cos 4\big(M+\sigma\big),$$
$$\vdots$$

and hence:

$$\sqrt{1+k^2\sin^2(M+\sigma)} = \left(1+\frac{k^2}{4} - \frac{3}{64}k^4+\cdots\right) + \left(-\frac{k^2}{4}+\frac{k^4}{16}+\cdots\right)\cos 2(M+\sigma) - \frac{k^4}{64}\cos 4(M+\sigma) + \cdots.$$

Substituting the above equation into (5.88) and integrating, with

$$\int_0^\sigma \cos 2(M+\sigma)\mathrm{d}\sigma = \frac{1}{2}\left[\sin 2(M+\sigma) - \sin 2M\right] = \sin\sigma\cos(2M+\sigma),$$
$$\int_0^\sigma \cos 4(M+\sigma)\mathrm{d}\sigma = \frac{1}{2}\sin 2\sigma\cos(4M+2\sigma),$$

one gets

$$S=b\left[\left(1+\frac{k^2}{4}-\frac{3}{64}k^4\right)\sigma-\left(\frac{k^2}{4}-\frac{k^4}{16}\right)\sin\sigma\cos(2M+\sigma)-\frac{k^4}{128}\sin 2\sigma\cos(4M+2\sigma)\right],$$

or

$$S = b[A\sigma - B\sin\sigma\cos(2M+\sigma) - C\sin 2\sigma\cos(4M+2\sigma)], \tag{5.89}$$

where

$$\left.\begin{aligned} A &= 1+\frac{k^2}{4}-\frac{3}{64}k^4 \\ B &= \frac{k^2}{4}-\frac{k^4}{16} \\ C &= \frac{k^4}{128} \end{aligned}\right\}.$$

The expressions for calculating σ from a given S can also be obtained from (5.89), as:

$$\sigma'' = \alpha S + \beta\sin\sigma\cos(2M+\sigma) + \gamma\sin 2\sigma\cos(4M+2\sigma), \tag{5.90}$$

where

$$\left.\begin{aligned} \alpha &= \frac{\rho''}{bA} = \frac{\rho''}{b}\left(1-\frac{k^2}{4}+\frac{7k^4}{64}\right) \\ \beta &= \frac{B\rho''}{A} = \rho''\left(\frac{k^2}{4}-\frac{k^4}{8}\right) \\ \gamma &= \frac{C\rho''}{A} = \rho''\left(\frac{k^4}{128}\right) \end{aligned}\right\}.$$

The m in k and the M in (5.89) and (5.90) can be determined by the right triangle $P_0'Q_1P_1'$ according to the expressions:

$$\left.\begin{aligned} \tan M &= \frac{\tan u_1}{\cos A_1} \\ \tan m &= \tan A_1\cos M \\ \sin m &= \cos u_1\sin A_1 \end{aligned}\right\}. \tag{5.91}$$

The Relationship Between l and λ

Integrating (5.86) gives the relationship between l and λ. As stated above, u has been converted into the function of σ, so if $\mathrm{d}\sigma$ is used to represent $\mathrm{d}\lambda$, the differential equation (5.86) can be solved.

It follows from (5.82) that:

$$\mathrm{d}\lambda = \frac{\sin A}{\cos u}\mathrm{d}\sigma.$$

Also, let P_2' be a moving point P'; from the right triangle $P_0'Q_2P_2'$ we have:

$$\sin A = \frac{\sin m}{\cos u},$$

and we obtain:

$$\mathrm{d}\lambda = \frac{\sin m}{\cos^2 u}\mathrm{d}\sigma.$$

Substituting the above equation into (5.86) yields:

$$\mathrm{d}l = \sqrt{1 - e^2\cos^2 u}\,\frac{\sin m}{\cos^2 u}\mathrm{d}\sigma.$$

In order to find the integral, we expand the above equation into the series of u, hence:

$$\begin{aligned}\mathrm{d}l &= \left(1 - \frac{e^2}{2}\cos^2 u - \frac{e^4}{8}\cos^4 u + \cdots\right)\frac{\sin m}{\cos^2 u}\mathrm{d}\sigma \\ &= \mathrm{d}\lambda - \sin m\left(\frac{e^2}{2} + \frac{e^4}{8}\cos^2 u + \cdots\right)\mathrm{d}\sigma.\end{aligned}$$

Substituting (5.87) into the above equation gives:

$$\begin{aligned}\mathrm{d}l &= \mathrm{d}\lambda - \sin m\left[\frac{e^2}{2} + \frac{e^4}{8} - \frac{e^4}{8}\cos^2 m\sin^2(M+\sigma) + \cdots\right]\mathrm{d}\sigma \\ &= \mathrm{d}\lambda - \sin m\left[\frac{e^2}{2} + \frac{e^4}{8} - \frac{e^4}{16}\cos^2 m + \frac{e^4}{16}\cos^2 m\cos 2(M+\sigma) + \cdots\right]\mathrm{d}\sigma.\end{aligned}$$

Integrating the above equation (accurate to the term of e^4) results in:

$$l = \lambda - \sin m\left[\left(\frac{e^2}{2} + \frac{e^4}{8} - \frac{e^4}{16}\cos^2 m\right)\sigma + \frac{e^4}{16}\cos^2 m \sin\sigma\cos(2M+\sigma)\right],$$

Or

$$l'' = \lambda'' - \sin m\left[\alpha' \sigma'' + \beta' \sin\sigma\cos(2M+\sigma)\right], \tag{5.92}$$

Where

$$\left.\begin{aligned} \alpha' &= \left(\frac{1}{2} + \frac{e^2}{8} - \frac{k'^2}{16}\right)e^2 \\ \beta' &= \frac{e^2}{16}k'^2\rho'' \\ k'^2 &= e^2\cos^2 m \end{aligned}\right\}.$$

Equations (5.89) or (5.90) and (5.92) are the formulae for projections of the side lengths and longitude differences in the Bessel's solution of geodetic problems. These expressions are expanded in the power series of e'^2 and e^2, not in the power series of S. The accuracy of the solution is related to the expanded terms rather than the distance. The formulae can be applied to the solution of the geodetic problem over any distances (particularly long distances). The main deficiency of Bessel's formula is that, while computing σ given S and computing λ given l, iteration is needed and the auxiliary quantities m and M should also be calculated beforehand.

5.6.4 *Computations of Bessel's Direct Solution of the Geodetic Problem*

Steps for Solution

Project the Ellipsoidal Elements onto a Spherical Surface

1. Given B_1, find u_1:

$$\tan u_1 = \sqrt{1-e^2}\tan B_1. \tag{5.93}$$

2. Calculate the auxiliary quantities m and M:

$$\left.\begin{aligned} \sin m &= \cos u_1 \sin A_1 \\ \tan M &= \frac{\tan u_1}{\cos A_1} \end{aligned}\right\}. \tag{5.94}$$

3. Convert S into σ:

$$\sigma = \alpha S + \beta \sin\sigma \cos(2M + \sigma) + \gamma \sin 2\sigma \cos(4M + 2\sigma). \tag{5.95}$$

The right-hand side of the above equation has the quantity σ that needs to be computed, hence iteration is needed.

The first-time approximation takes:

$$\sigma_0 = \alpha S.$$

The approximation of i times is:

$$\sigma_i = \alpha S + \beta \sin\sigma_{i-1} \cos(2M + \sigma_{i-1}) + \gamma \sin 2\sigma_{i-1} \cos(4M + 2\sigma_{i-1}),$$

until the required accuracy is satisfied. For instance, if we want $\Delta S < 0.3$ m, then we need $|\sigma_i - \sigma_{i\_1}| < 0.01''$ (i.e., 2.8×10^{-6}); if we want $\Delta S < 0.03$ m, then we need $|\sigma_i - \sigma_{i\_1}| < 0.001''$ (i.e., 2.8×10^{-7}).

For precise solutions, α, β, and γ are calculated according to the expressions below:

$$\left.\begin{aligned} \alpha &= \frac{\rho\sqrt{1+e'^2}}{a}\left(1 - \frac{k^2}{4} + \frac{7k^4}{64} - \frac{15k^6}{256}\right) \\ \beta &= \rho\left(\frac{k^2}{4} - \frac{k^4}{8} + \frac{37k^6}{512}\right) \\ \gamma &= \rho\left(\frac{k^4}{128} - \frac{k^6}{128}\right) \\ k^2 &= e'^2 \cos^2 m \end{aligned}\right\}.$$

For approximate solutions in meters, the above equations are accurate to the term of k^4, while for approximate solutions in hectometers, the γ term can also be neglected.

Solution of Spherical Triangles

1. Find A_2

   Since $A_2 = A'_2 + 180^\circ$, from the right-angled spherical triangle $P_0'Q_2P_2'$, one obtains:

$$\tan A_2 = \tan A_2' = \frac{\tan m}{\cos(M+\sigma)}. \tag{5.96}$$

2. Find u_2

   Again from the right-angled spherical triangle $P_0'Q_2P_2'$, we have:

$$\tan u_2 = \cos A_2' \tan(M+\sigma) = -\cos A_2 \tan(M+\sigma). \tag{5.97}$$

3. Compute λ

   From the right-angled spherical triangles $P_0'Q_1P_1'$ and $P_0'Q_2P_2'$, we get:

$$\left.\begin{aligned} \tan\lambda_1 &= \sin m \tan M = \sin u_1 \tan A_1 \\ \tan\lambda_2 &= \sin m \tan(M+\sigma) = \sin u_2 \tan A_2 \\ \lambda &= \lambda_2 - \lambda_1 \end{aligned}\right\}. \tag{5.98}$$

   So far, the three unknowns A_2, u_2, and λ on the spherical surface have all been obtained.

Project Elements from the Spherical Surface onto the Ellipsoidal Surface

1. Given u_2, find B_2:

$$\tan B_2 = \sqrt{1+e'^2}\tan u_2. \tag{5.99}$$

2. Convert λ into l and find L_2:

$$l = \lambda - \sin m\left[\alpha'\sigma + \beta' \sin\sigma\cos(2M+\sigma) + \gamma' \sin 2\sigma \cos(4M+2\sigma)\right], \tag{5.100}$$

   where

$$\left.\begin{aligned}
\alpha' &= \left(\frac{e^2}{2}+\frac{e^4}{8}+\frac{e^6}{16}\right) - \frac{e^2}{16}\left(1+e^2\right)k'^2 + \frac{3}{128}e^2k'^4 \\
\beta' &= \rho\left[\frac{e^2}{16}\left(1+e^2\right)k'^2 - \frac{e^2}{32}k'^4\right] \\
\gamma' &= \rho\frac{e^2}{256}k'^4 \\
k'^2 &= e^2\cos^2 m
\end{aligned}\right\},$$

The maximum value of γ' is 0.0002″; hence, the γ' term in (5.100) is generally negligible.

For the approximate solutions in meters, α' and β' can be calculated according to (5.92); for the approximate solutions in hectometers, one can calculate according to equations that leave out the term of k'^2:

$$\alpha' = \left(\frac{1}{2}+\frac{e^2}{8}\right)e^2.$$

Finally, one obtains:

$$L_2 = L_1 + l \tag{5.101}$$

Determination of the Quadrant

With the above formulae, m, M, λ_1, λ_2, and A_2 are calculated using trigonometric functions. Therefore, it is still necessary to discuss the determination of their quadrants.

To determine quadrants for these quantities easily, we draw Figs. 5.44, 5.45, 5.46, and 5.47, which denote the quadrants in which m, M, λ_1, λ_2, and A_2 lie when point P_1' is in the northern hemisphere, i.e., u_1 is positive, and A_1 lies in quadrants I, II, III, and IV. In each figure, P_2 is assumed to have three positions, denoted by P_2', P_2'', and P_2''', respectively. Correspondingly, λ_2 lies in three different quadrants. In the figures, the great circle intersects the equator at P_0' and P_0'', and dashed lines denote the back side of the sphere. Figure 5.44 demonstrates that A_1 is in quadrant I, and here m, M, and λ_1 are all in quadrant I. At points P_2', λ_2 and A_2' are both in the quadrant I; at points P_2'', λ_2 and A'_2 are both in quadrant II; at points P_2''', λ_2 lies in quadrant III while A_2' lies in quadrant II. Since $A_2 = A'_2 \pm 180°$, A_2 lies in quadrant III or quadrant IV. Similarly, we can also explain the situations in Figs. 5.45, 5.46, and 5.47.

Fig. 5.44 shows that λ_2 and $(M + \sigma)$ are in the same quadrant. The quadrant of A_2 can be determined by the sign of tan A_2: when tan A_2 is positive, A_2 lies in quadrant

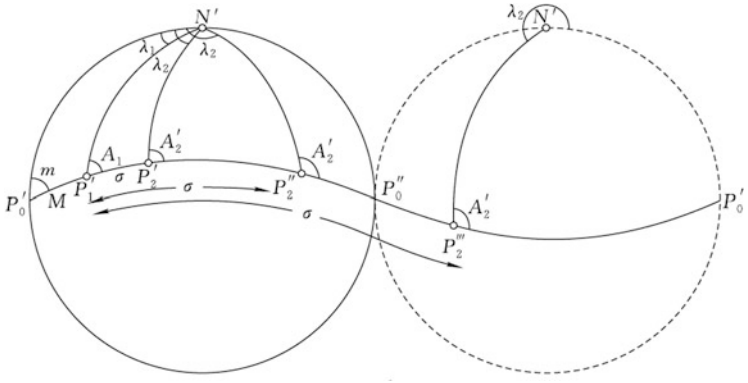


Fig. 5.44 A_1 lies in quadrant I (m, M, and λ_1 are all in quadrant I, λ_2 and $M + \sigma$ are in the same quadrant, A_2 is in quadrant IV or III). *Dashed line* indicates the back side of the sphere

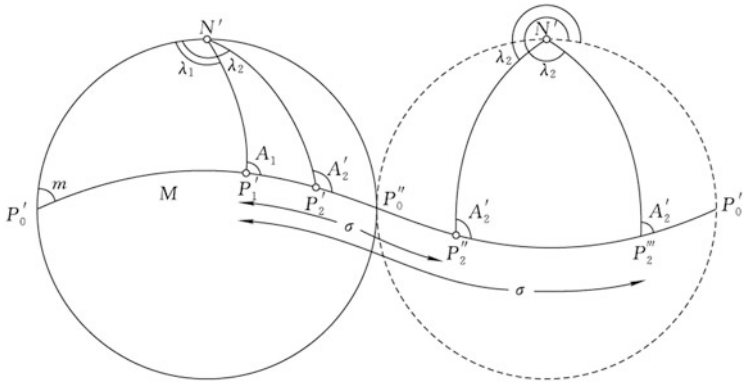


Fig. 5.45 A_1 lies in quadrant II (m in quadrant I, M and λ_1 in quadrant II, λ_2 and $M + \sigma$ are in the same quadrant, A_2 is in quadrant IV or III). *Dashed line* indicates the back side of the sphere

III; when $\tan A_2$ is negative, A_2 lies in quadrant IV. The above conclusion also applies to the condition where A_1 lies in quadrant II (Fig. 5.45).

However, when A_1 lies in quadrants III and IV (Figs. 5.46 and 5.47), λ_2 is in the same quadrant as $360^\circ - (M + \sigma)$. When $\tan A_2$ is positive, A_2 is in quadrant I; when $\tan A_2$ is negative, A_2 is in quadrant II.

From the above four figures, it is not difficult to obtain the quadrant table of m, M, λ_1, λ_2, and A_2 when u_1 is positive (in the northern hemisphere); see Table 5.9. When u_1 is negative (in the southern hemisphere), we obtain another table, which is omitted here.

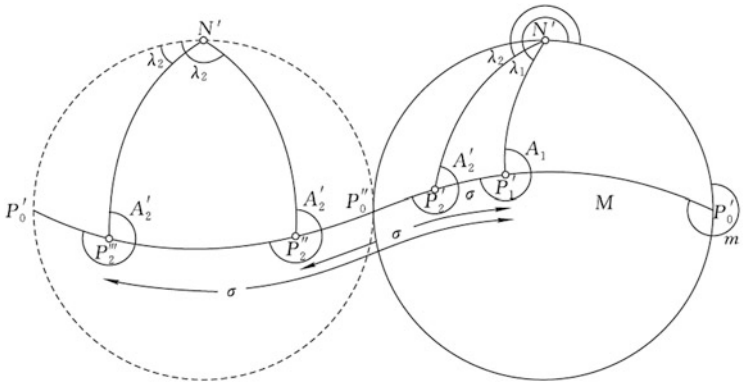


Fig. 5.46 A_1 lies in quadrant III (m in quadrant IV, M in quadrant II, and λ_1 lies in quadrant III, λ_2 and $360^{\circ} - (M + \sigma)$ are in the same quadrant, and A_2 lies in quadrant I or II). *Dashed line* indicates the back side of the sphere

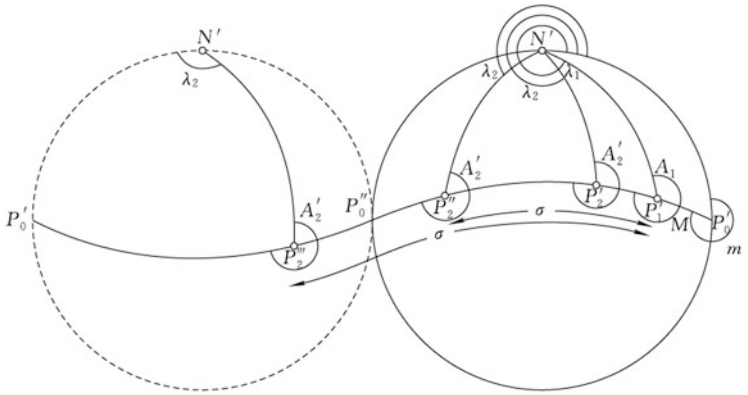


Fig. 5.47 A_1 lies in quadrant IV (m in quadrant IV, M in quadrant I, λ_1 lies in quadrant IV, λ_2 and $360^{\circ} - (M + \sigma)$ are in the same quadrant, and A_2 lies in quadrant II or I). *Dashed line* indicates the back side of the sphere

Table 5.9 Determination of the quadrant of the azimuths (A_1 and A_2), longitudes (λ_1 and λ_1), and auxiliary quantities (m and M) in the solution of the geodetic problem ($u_1 > 0$)

A_1	m	M	λ_1	λ_2	A_2
I	I	I	I	Same quadrant with $(M + \sigma)$	tan A_2 positive, in quadrant III
II	I	II	II		tan A_2 negative, in quadrant IV
III	IV	II	III	Same quadrant with $360^{\circ} - (M + \sigma)$	tan A_2 positive, in quadrant I
IV	IV	I	IV		tan A_2 negative, in quadrant II

Example

The solutions of long-distance geodetic problems are vital and useful in navigation and long-distance missile launching. We hereby provide the block diagram and instance of computation for approximate solutions in meters. Likewise, we can also program the block diagram for approximate solutions in hectometers and produce precise solutions (precise solutions or approximate solutions in hectometers) according to the formulae.

Block Diagram

The block diagram of computation for the solution of the direct geodetic problem is shown in Fig. 5.48.

Computations

A sample computation for the solution of the direct geodetic problem is provided in Table 5.10.

5.6.5 *Computations of Bessel's Inverse Solution of the Geodetic Problem*

Steps for Solution

Project the Ellipsoidal Elements onto the Spherical Surface

1. Given B, find u
   u can be obtained from:

$$\left.\begin{aligned}\tan u_1 &= \sqrt{1-e^2}\tan B_1\\ \tan u_2 &= \sqrt{1-e^2}\tan B_2\end{aligned}\right\}. \tag{5.102}$$

2. Given l, find λ
   In the inverse solution, given the longitude difference l on the ellipsoid, the corresponding longitude difference λ on the spherical surface is still unknown. To compute λ given l, from (5.92), obviously we first need to calculate σ, m, and M since all of them are involved in the reckoning of the correction terms. The accuracy required is not high and generally approximations made twice will be satisfactory.

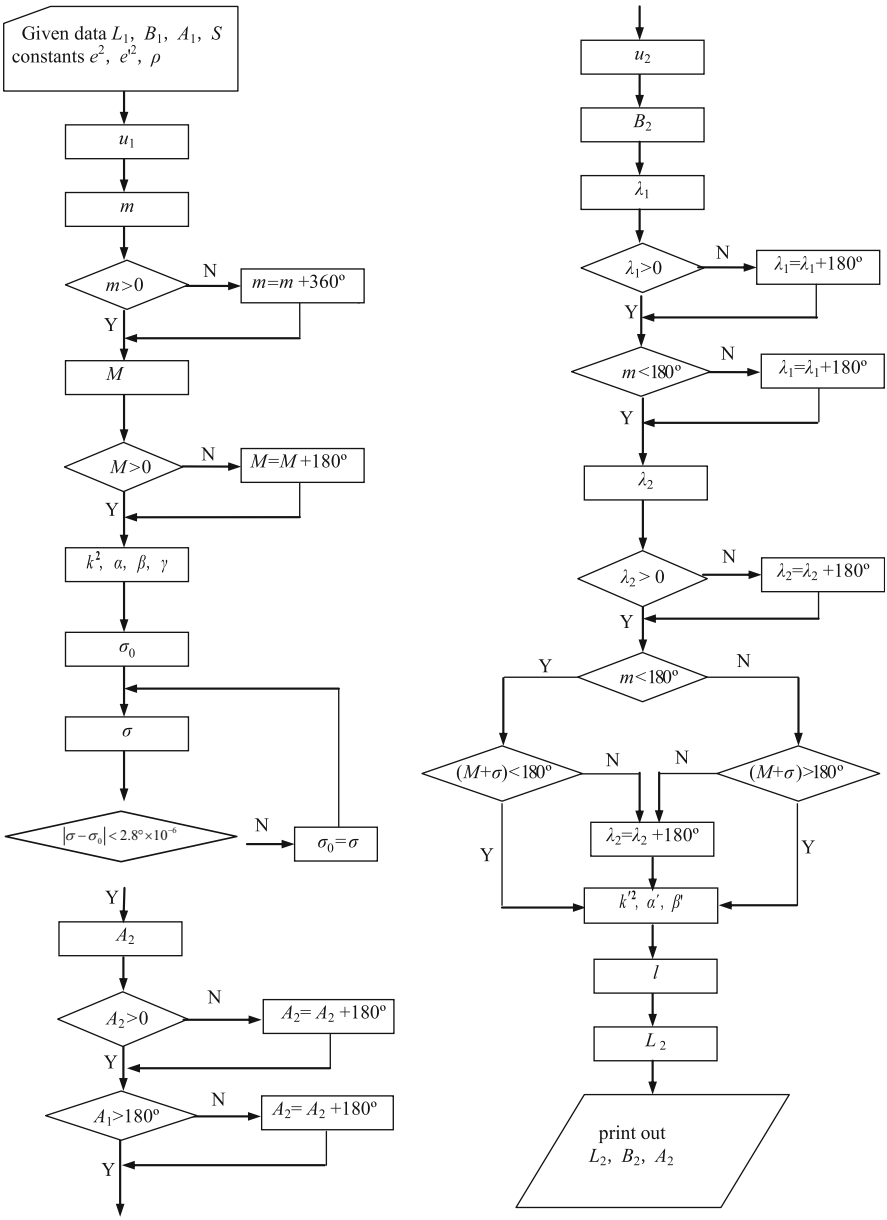


Fig. 5.48 Block diagram of the solution of direct geodetic problem (Y denotes yes, N denotes no)

Table 5.10 Sample computation for the direct solution of the geodetic problem

Given data	Ellipsoidal parameters	Computational results
$L_1 = 90^\circ\ 00'\ 00.11''$	Krassowski Ellipsoid	$L_2 = 215^\circ\ 59'\ 04.333''$
$B_1 = 35^\circ\ 00'\ 00.22''$		$B_2 = -30^\circ\ 29'\ 20.964''$
$A_1 = 100^\circ\ 00'\ 00.33''$		$A_2 = 290^\circ\ 32'\ 53.389''$
$S = 15\ 000\ 000.2$ m	GRS75 Ellipsoid	$L_2 = 215^\circ\ 59'\ 13.059''$
		$B_2 = -30^\circ\ 29'\ 23.867''$
		$A_2 = 290^\circ\ 32'\ 48.833''$
	GRS80 Ellipsoid	$L_2 = 215^\circ\ 59'\ 13.306''$
		$B_2 = -30^\circ\ 29'\ 23.947''$
		$A_2 = 290^\circ\ 32'\ 48.708''$

Applying the cosine theorem to the triangle $P_1'P_2'N'$ in Fig. 5.43 gives (l is used to approximately replace λ, and hence the σ obtained is an estimated value σ_0):

$$\cos\sigma_0 = \sin u_1 \sin u_2 + \cos u_1 \cos u_2 \cos l. \tag{5.103}$$

From triangles $P_1'P_2'Q_1$ and $P_1'P_2'N'$, one obtains:

$$\sin m_0 = \cos u_1 \sin A_1 = \cos u_1 \cos u_2 \frac{\sin l}{\sin \sigma_0}. \tag{5.104}$$

The approximate estimated value m_0 of m is hereby obtained.
With (5.92), we take the approximate correction for λ:

$$\Delta\lambda = \lambda - l \approx \alpha' \sigma_0 \sin m_0 = 0.003351\sigma_0 \sin m_0. \tag{5.105}$$

Hence, we have:

$$\lambda_0 = l + \Delta\lambda.$$

Taking (5.103) and differentiating σ and l results in:

$$-\sin\sigma_0 \Delta\sigma = -\cos u_1 \cos u_2 \sin\lambda_0 \Delta\lambda,$$

Or

$$\Delta\sigma = \cos u_1 \cos u_2 \frac{\sin\lambda_0}{\sin\sigma_0}\Delta\lambda = \sin m_0 \Delta\lambda. \tag{5.106}$$

Hence, the result is:

$$\sigma_1 = \sigma_0 + \Delta\sigma.$$

Substituting λ_0 and σ_1 into (5.104) produces:

$$\sin m = \cos u_1 \cos u_2 \frac{\sin \lambda_0}{\sin \sigma_1}. \tag{5.107}$$

In addition, applying the cotangent theorem to the triangle $P_1' P_2' N'$, replacing λ with λ_0 gives:

$$\left.\begin{aligned} \cot A_1^{\circ} &= \tan u_2 \cos u_1 \csc \lambda_0 - \sin u_1 \cot \lambda_0 \\ \tan A_1^{\circ} &= \frac{\sin \lambda_0}{\cos u_1 \tan u_2 - \sin u_1 \cos \lambda_0} \end{aligned}\right\}. \tag{5.108}$$

In this way the estimated value A_1° of A_1 is obtained.

In the right triangle $P_1' P_2' Q_1$, we have:

$$\left.\begin{aligned} \cot M &= \frac{\sin m \cot A_1{}^{\circ}}{\sin u_1} \\ \tan M &= \frac{\sin u_1}{\sin m} \tan A_1{}^{\circ} \end{aligned}\right\}. \tag{5.109}$$

Hence, the estimated value of M is obtained.

We calculate m according to (5.107) and compute α' and β' according to (5.100) and (5.92); the required accuracy is the same as for the direct solution. Finally, we calculate the longitude difference on the spheroid using the expression:

$$\lambda = l + \sin m \left[\alpha' \sigma + \beta' \sin \sigma \cos (2M + \sigma) \right]. \tag{5.110}$$

Solution of Spherical Triangles

1. Find σ

   As shown in Fig. 5.43, applying the cosine law to the spherical triangle $P_1' P_2' N'$ gives:

$$\cos \sigma = \sin u_1 \sin u_2 + \cos u_1 \cos u_2 \cos \lambda. \tag{5.111}$$

2. Find A_1 and A_2

   As shown in Fig. 5.43, applying the cotangent law to the spherical triangle $P_1' P_2' N'$ gives:

$$\left.\begin{aligned}\cot A_1 &= \tan u_2 \cos u_1 \csc\lambda - \sin u_1 \cot\lambda \\ \tan A_1 &= \frac{\sin\lambda}{\cos u_1 \tan u_2 - \sin u_1 \cos\lambda}\end{aligned}\right\}, \tag{5.112}$$

$$\left.\begin{aligned}\cot A_2 &= \sin u_2 \cot\lambda - \tan u_1 \cos u_2 \csc\lambda \\ \tan A_2 &= \frac{\sin\lambda}{\sin u_2 \cos\lambda - \tan u_1 \cos u_2}\end{aligned}\right\}. \tag{5.113}$$

Convert the Spherical Elements into Ellipsoidal Elements

1. Geodetic azimuths A_1 and A_2 remain unconverted.
2. Convert σ into S

   S is given by:

$$S = \frac{1}{\alpha}[\sigma - \beta \sin\sigma \cos(2M + \sigma) - \gamma \sin 2\sigma \cos(4M + 2\sigma)], \tag{5.114}$$

where α, β, and γ are calculated according to the precise solution formula. The accuracy required is the same as in the direct solution.

Determination of the Quadrant

In computation we also need to determine the quadrants of A_1 and A_2. No more details will be given here; see Table 5.9.

Example

Block Diagram

The block diagram of computation for the solution of the inverse geodetic problem is given in Fig. 5.49.

Computations

Table 5.11 provides a sample computation for the solution of the inverse geodetic problem.

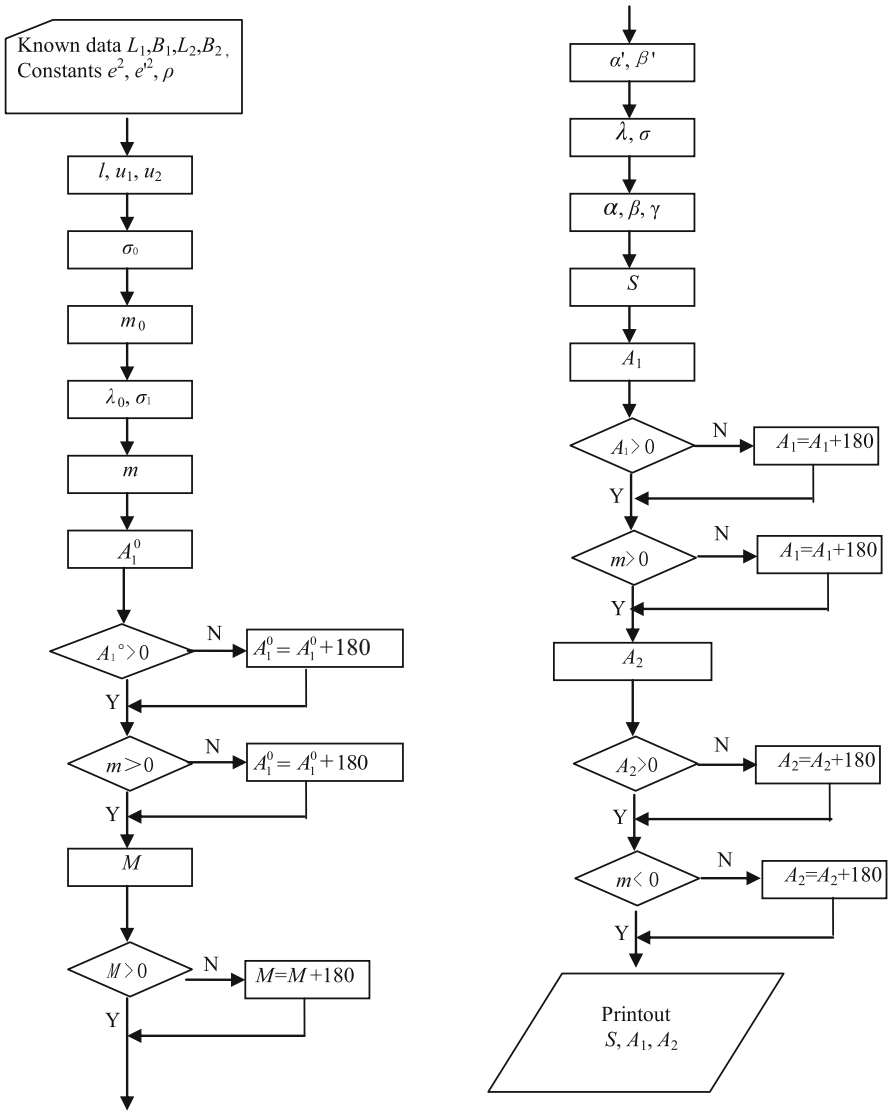


Fig. 5.49 Block diagram of the inverse solution of the geodetic problem (where Y denotes Yes, N denotes No)

Table 5.11 Sample computation for the inverse solution of the geodetic problem

Given data	Ellipsoidal parameters	Computational results
$L_1 = 90^\circ\ 00'\ 00.11''$	Krassowski Ellipsoid	$S = 15\ 000\ 000.330$ m
$B_1 = 35^\circ\ 00'\ 00.22''$		$A_1 = 100^\circ\ 00'\ 00.321''$
$L_2 = 215^\circ\ 59'\ 04.34''$		$A_2 = 290^\circ\ 32'\ 53.392''$
$B_2 = -30^\circ\ 29'\ 20.96''$	GRS75 Ellipsoid	$S = 14\ 999\ 751.047$ m
		$A_1 = 100^\circ\ 00'\ 00.228''$
		$A_2 = 290^\circ\ 32'\ 53.329''$
	GRS80 Ellipsoid	$S = 14\ 999\ 744.004$ m
		$A_1 = 100^\circ\ 00'\ 00.229''$
		$A_2 = 290^\circ\ 32'\ 53.329''$

Review and Study Questions

1. Why do we choose a reference ellipsoid as the reference surface on which geodetic surveying computations are performed? What are the requirements for the reference surface selected for geodetic computations?
2. Explain the concepts of the Earth ellipsoid and the reference ellipsoid.
3. Given the flattening α and eccentricity e, can we define the shape and size of the reference ellipsoid and why?
4. What is the difference between astronomical longitude and latitude and the geodetic longitude and latitude?
5. Construct a sketch to derive the formulae for transformations between L, B, H and X, Y, Z.
6. Construct and derive the formulae for transformations between the geocentric coordinate system *O-XYZ* and the topocentric coordinate system *P-xyz* (list the formulae that include rotation matrices and translation vectors will do).
7. Which point on the ellipsoid has an R_A independent of directions? Please illustrate.
8. Write out a set of formulae for the radius of curvature of the meridian, the radius of curvature in the prime vertical, and the mean radius of curvature, using V, c and W, a, and specify the relationship between the values of their results.
9. Analyze the laws of changes of the meridional radius of curvature and radius of curvature in the prime vertical with the changes of latitude (trajectory of the center of curvature when $0 \leq B \leq 90^\circ$).
10. Construct and derive the integral of the meridian arc length.
11. Construct and derive the formula for the length of a parallel arc.
12. Why does the length of the meridian arc of the unit difference in latitude in the northern hemisphere turn out to be longer in the north and shorter in the south? Why is the arc length along a parallel of the unit difference in longitude shorter in the north and longer in the south?

13. In astro-geodetic surveying, the mean square errors of the astronomical determinations of latitude and longitude are respectively $\pm 0.3''$ and ± 0.02 s (equivalent to $\pm 0.3''$). So if $B = 45°$, what are the arc lengths along an ellipsoidal meridian and along a parallel and what are these mean square errors equivalent to?
14. Why can we transform the solution of ellipsoidal triangles to the solution of plane triangles? Please explain Legendre's theorem by which ellipsoidal triangles are solved.
15. What are the reasons for the formation of reciprocal normal sections? Please describe the rules of position of the normal section and the reverse normal section. Under what condition do the reciprocal normal sections between two points on the ellipsoid coincide?
16. What is the geodesic? Please describe the relationship between positions on the geodesic and reciprocal normal sections.
17. Which normal sections on the ellipsoid are geodesics? Are parallel and prime vertical geodesics?
18. Derive the differential equations for geodesics.
19. Derive Clairaut's equation for geodesics and illustrate the meaning of this equation.
20. Given the formula for correction of the vertical deflection:

$$\delta_1 = R - R_1 = -(\xi \sin A - \eta \cos A) \cot z_1,$$

mark out the quantities δ_1, ξ, η, A, and Z_1 on the auxiliary sphere.
21. Construct and explain the meaning of skew normal correction.
22. Draw a diagram of correction for skew normals with the azimuth values of sides lying in the I, II, III, and IV quadrants, respectively and specify the sign of the correction for skew normals δ_2.
23. Explain the meaning of correction from normal section to geodesic.
24. Account for the meaning of reduction of the observed zenith distance.
25. What are the data that should be given to calculate the three corrections, i.e., correction for deflection of the vertical, correction for skew normals, and correction from normal section to geodesic? Under what conditions do the three corrections equal zero, respectively.
26. Account for the basic process of slope distance reduction.
27. Derive the approximate formula for the reduction of slope distance:

$$S = R_A \sigma,$$

where σ is given by $D^2 = (R_A + H_1)^2 + (R_A + H_2)^2 - 2(R_A + H_1)(R_A + H_2) \cos \sigma$.
28. What is gravimetric deflection of the vertical and astro-geodetic deflection of the vertical? What are the two factors that affect the value of deflection of the vertical?

29. Write out the formula for deflection of the vertical. What are the double-parallel conditions for the validity of this formula?
30. Explain the meaning of astronomical azimuth reduction.
31. What is the function of Laplace azimuth in the triangulateration network?
32. Explain the concepts of solutions of the direct and inverse geodetic problems.
33. Explain the basic principles of the point-by-point integration for solutions of direct geodetic problems.
34. Explain the steps and projection conditions of Bessel's solution of geodetic problems.
35. Derive the relationship between the reduced latitude and the geodetic latitude.
36. Prove that in Bessel's solution of the geodetic problem, the azimuth of a geodesic remains unchanged after projection.

Chapter 6
Gauss and UTM Conformal Projections and the Plane Rectangular Coordinate System

In Chap. 5 we established the relationship between geodetic elements on the Earth's surface and those on the ellipsoid. This chapter further establishes the corresponding relationship between geodetic elements on the ellipsoid and those on the plane. We will discuss the corresponding relationship between geodetic coordinates, geodesic direction, geodesic distance, and geodetic azimuth, and their corresponding counterparts on the plane. Such corresponding relationship is realized through mathematical projection methods, of which there are many. This chapter, however, is primarily concerned with the two conformal (orthomorphic) projections used in geodetic survey, i.e., the Gauss projection and the Universal Transverse Mercator (UTM) projection and establishes the projection relationship between the geodetic coordinate system and plane coordinate system, as well as the relationship between the geodetic control network on the ellipsoid and that on the plane.

6.1 Overview of Projection

6.1.1 Aims of Projection

The reference ellipsoid is the datum for geodetic computations (of geodetic coordinates, geodetic azimuth, and geodesic distance, etc.) and for study of the shape and size of the Earth (computations of vertical deflection and height anomaly). Geodetic coordinates on the ellipsoid are the fundamentals for geodetic survey. One of the roles of geodetic survey is to determine the coordinates of surface points to control topographic mapping. Maps are flat, so the coordinates of geodetic points used to control mapping have to be plane coordinates, otherwise they will be unrelated, for one belongs to the plane system and the other the ellipsoid system. Establishing the corresponding relationship between geodetic coordinates and plane coordinates therefore becomes necessary, which is called projection. Within

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_6,

a small area, the Earth's surface can be considered a plane, but for a large area it is subject to distortion if represented as a plane. Projection is essentially to establish functional models to allocate such distortions in a reasonable way.

In addition, from Bessel's formula for the solution of the geodetic problem in Chap. 5, we know that, although the ellipsoid is a mathematical surface, geodetic computations carried out on it are complicated. If elements on the ellipsoid are reduced to a plane, geodetic computations will become much simpler. For instance, coordinates of points on a plane can be solved according to simple formulae in plane trigonometry.

Thus, projection from the ellipsoid surface onto a plane is used to control topographic mapping and to simplify survey computations. The specialized discipline concerning problems of projection is called map projection. In geodetic survey, the projection of a geodetic control network is also studied in addition to the projection of coordinates for positions of points.

6.1.2 Definition of Projection

Projection means establishing a one-to-one correspondence between the geodetic elements on the ellipsoid and the corresponding elements on the plane according to certain mathematical rules. The former includes geodetic coordinates, geodesic direction, distance of geodesic line, geodetic azimuth, etc. Once the relationship between the coordinates of point positions is specified, the corresponding relationship between other elements will be determined as well; hence, the key to determining the projection relationship is to determine the projection relationship between the coordinates of point positions.

The "certain mathematical rules" mentioned can be expressed by the following equations:

$$\begin{cases} x = F_1(B, L) \\ y = F_2(B, L) \end{cases} \tag{6.1}$$

where (B, L) are the geodetic coordinates of a point on the ellipsoid, and (x, y) are the rectangular coordinates of this point after being projected onto the plane. It is obvious that (6.1) is single-valued, finite, and continuous.

Equation (6.1) expresses the analytical projection relationship between points on the ellipsoid and their corresponding points on the projection plane, without any geometrical meaning. Different projections actually determine the functional forms of F_1 and F_2 in the formulae according to their specified conditions. The Gauss projection (and the UTM projection) has its own particular conditions. Once F_1 and F_2 are determined, the geodetic coordinates of every point on the ellipsoid and the rectangular coordinates of their respective corresponding points can be determined accordingly.

How can F_1 and F_2 be determined? Projection problems have different solutions based on different demands. We turn these demands into certain mathematical conditions and allow them to be expressed in the projection formula to get the specific mathematical relationship. Next we will discuss the requirements of projections for controlling mapping.

6.1.3 Conformal Projection and Conformality

Projection means the planar representation of quantities on the ellipsoid, which will inevitably create distortions. A projection distortion refers to the alteration of angle, distance, or area after projection. Cones and cylinders are developable surfaces, on which quantities will not be distorted if represented on a plane. An ellipsoid and a sphere are undevelopable surfaces that crumple or fold when unrolled and flattened by force. Projection distortions are no doubt unfavorable, but they can be allocated and controlled reasonably by determining F_1 and F_2 in (6.1).

There are three kinds of projection distortions, i.e., angle distortion, distance distortion, and area distortion. They can be controlled according to specific needs. A distortion of some kind can be zero, such as when the projection is equiangular, equivalent (equal-area), or equidistant; distortions of all kinds can also coexist but have to be kept within a proper degree. It is obviously impossible to eliminate all distortions at the same time since the ellipsoid is an undevelopable surface and a projection will invariably introduce distortions.

For large-scale mapping, if figures on a map can be maintained conformable to the original on the ellipsoid within a certain area, i.e., angles remain undistorted after projection, then on such maps terrains and land features will be completely conformable to the real features, which will bring great convenience in use. A projection in which all angles at any point are preserved is known as the equiangular or conformal projection.

As shown in Fig. 6.1, a small midpoint polygon $OABCDE$ on the ellipsoid is conformally projected onto the plane as $O'A'B'C'D'E'$. Each line segment on the ellipsoid in Fig. 6.1 is a differential line segment (called an arc element), which is considered a straight line and remains the same after being projected onto the plane. According to the definition of conformal projection, the internal angles of every triangle are not altered after projection. Triangles are correctly represented after conformal projection, so their corresponding sides are proportional, thus:

$$\frac{O'A'}{OA} = \frac{O'B'}{OB} = \frac{O'C'}{OC} = \frac{O'D'}{OD} = \frac{O'E'}{OE} = m = \text{constant}$$

where m is the scale factor. Consequently, for certain points in conformal projection, the scale factor m is independent of direction. However, this property of conformal projection is conditional and is only valid within a small area. It is impossible that a map projection can be achieved in which large areas are rendered

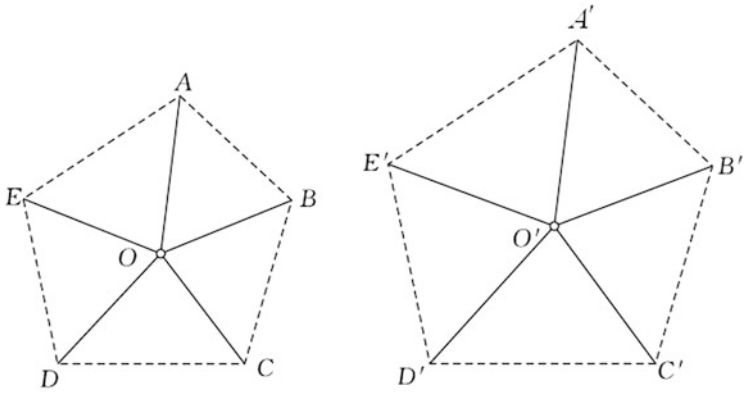


Fig. 6.1 A figure correctly represented by conformal projection

with true shape, i.e., that the ellipsoid can be unfolded or unrolled onto the plane without distortions. Hence, within a large area, the scale factor m varies from point to point, i.e., m is dependent on the position of points. To sum up, the conformality of conformal projection is that, in conformal projection, the scale factor m is independent of direction but dependent on the position of points.

Map projections are of many kinds. Taking into account the nature of distortions, apart from the aforementioned conformal projection, there are also equidistant projections (the distances between any arbitrary two points on the ellipsoid remain unchanged after projection onto the plane) and equivalent projections (the area on the ellipsoid is preserved after projection onto the plane).

Conformality is the unique property that differentiates the conformal projection from other projections. The general conditions for a conformal projection are based on this property.

6.2 General Condition for Conformal Projection

6.2.1 Overview

Conformal projection is one kind of map projection, and the Gauss projection and UTM projection are two kinds of conformal projection. Therefore, the conformal projection needs to be studied prior to study concerning the Gauss projection or UTM projection. The task of this section is to derive the general condition for conformal projection, in combination with the particular conditions for Gauss and UTM projections; the formula for the Gauss projection or UTM projection will then be derived.

To derive the general condition for the conformal projection, we have to grasp its unique property that differs it other projections, i.e., in conformal projection the

scale factor is independent of direction. This is the fundamental idea in deducing the general condition for conformal projection. The specific steps are:

1. Based on definition of the scale factor:

$$m = \frac{\mathrm{d}s}{\mathrm{d}S},$$

   where ds is the arc element on the projection plane and dS is the arc element on the ellipsoid; write down its specific expression.
2. We deduce the general condition for conformal projection according to the fact that the scale factor m is independent of the azimuth A.

6.2.2 Expression of Scale Factor

In Fig. 6.2, the left part of the diagram is the ellipsoid surface and the right part is the projection plane. We establish the Cartesian coordinate system on the projection plane with o as the origin, the x-axis as the vertical axis, and the y-axis as the horizontal axis. The coordinate system so defined is a left-hand system, while in most cases the plane analytical system is displayed as the right-hand system, because the azimuth defined in geodesy is the angle measured clockwise from the north direction (which is the same as the angle defined in the left-hand system). In many books, the plane coordinates are expressed as a distance in meters to the east, referred to as the "Easting," and a distance in meters to the north, referred to as the "Northing" (FGDC 2001; NIMA 1989, 1990; Maling 1992; DeMers 2005; Galati 2006).

There are two infinitely near points P and P_1 on the ellipsoid surface, which are projected as P' and $P_1{}'$. Their coordinates are shown in Fig. 6.2. dS is the geodesic arc element with an azimuth A and the projected arc element on the plane is ds. We draw the meridian and parallel through points P and P_1, respectively, from the differential triangle PP_1P_2 according to the differential formulae for the arc lengths of the meridian and parallel, and the expression of the arc element dS on the ellipsoid can be written as:

$$\mathrm{d}S^2 = M^2\mathrm{d}B^2 + r^2\mathrm{d}L^2 = r^2\left[\frac{M^2}{r^2}(\mathrm{d}B)^2 + (\mathrm{d}L)^2\right], \tag{6.2}$$

where the curvature radii M and r are functions of the latitude B. To simplify the derivation process of the formula, let:

$$\mathrm{d}q = \frac{M}{r}\mathrm{d}B, \tag{6.3}$$

then (6.2) can be simplified as:

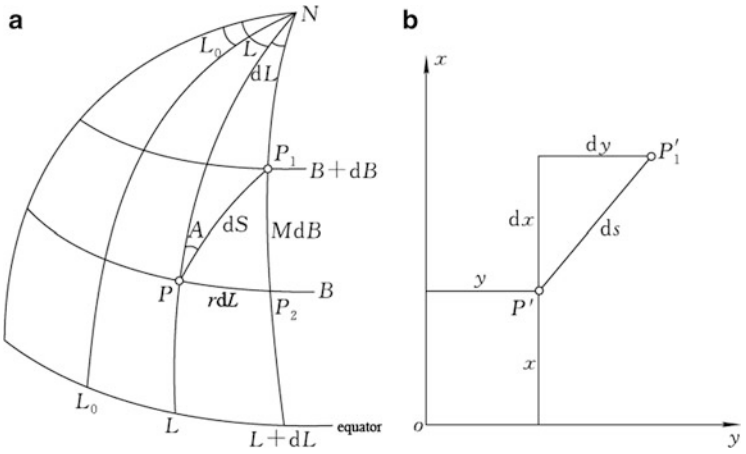


Fig. 6.2 Scale factor $m = \frac{ds}{dS}$ for (**a**) an ellipsoid surface and (**b**) a projection plane

$$dS^2 = r^2\left[(dq)^2 + (dL)^2\right]. \tag{6.4}$$

From (6.3) we know that q is merely the function of B, which can be called isometric latitude. The relational expression between the geodetic latitude B and isometric latitude q can be determined according to:

$$q = \int \frac{M}{r} dB,$$

which means that, given a geodetic latitude, one corresponding isometric latitude can be obtained accordingly. Isometric latitude, however, is meaningless in practical use. There is no reason to calculate the isometric latitude and it is introduced here only for the convenience of formula derivation.

The arc element ds on the projection plane can be written directly according to the formula for an arc element on the plane curve:

$$ds^2 = (dx)^2 + (dy)^2. \tag{6.5}$$

Then we have:

$$m^2 = \left(\frac{ds}{dS}\right)^2 = \frac{(dx)^2 + (dy)^2}{r^2\left[(dq)^2 + (dL)^2\right]}. \tag{6.6}$$

The deduction of a conformal condition has to be based on the fact that the scale factor m is independent of the azimuth A. Thus, the azimuth has to be introduced in (6.6) to change the above formula. We know that projection means to determine

specifically the functions F_1 and F_2 of (6.1), i.e., to establish the functional relationship between the plane coordinates (x, y) and the geodetic coordinates (L, B). Since the geodetic latitude B is related to q, the projection problem is to establish the functional relationship between (x, y) and (L, q). The geodetic longitude L of a point is referenced to the initial geodetic meridian. If another meridian L_0 is instead referred to rather than the initial geodetic meridian, then L should be transformed into the longitude difference l, where $l = L - L_0$, and $\mathrm{d}l = \mathrm{d}L$. Therefore, establishing the relationship between (x, y) and (L, q) will be changed into that between (x, y) and (l, q). We assume their relationship is:

$$\left.\begin{aligned} x &= f_1(q,l) \\ y &= f_2(q,l) \end{aligned}\right\}. \tag{6.7}$$

The total differential of the above equation gives:

$$\left.\begin{aligned} \mathrm{d}x &= \frac{\partial x}{\partial q}\mathrm{d}q + \frac{\partial x}{\partial l}\mathrm{d}l \\ \mathrm{d}y &= \frac{\partial y}{\partial q}\mathrm{d}q + \frac{\partial y}{\partial l}\mathrm{d}l \end{aligned}\right\}.$$

Substituting the above equations into (6.5) yields:

$$\mathrm{d}s^2 = \left[\frac{\partial x}{\partial q}\mathrm{d}q + \frac{\partial x}{\partial l}\mathrm{d}l\right]^2 + \left[\frac{\partial y}{\partial q}\mathrm{d}q + \frac{\partial y}{\partial l}\mathrm{d}l\right]^2$$

$$= \left[\left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2\right](\mathrm{d}q)^2 + 2\left[\frac{\partial x}{\partial q}\frac{\partial x}{\partial l} + \frac{\partial y}{\partial q}\frac{\partial y}{\partial l}\right]\mathrm{d}q\cdot\mathrm{d}l + \left[\left(\frac{\partial x}{\partial l}\right)^2 + \left(\frac{\partial y}{\partial l}\right)^2\right](\mathrm{d}l)^2.$$

With

$$\left.\begin{aligned} E &= \left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2 \\ F &= \frac{\partial x}{\partial q}\frac{\partial x}{\partial l} + \frac{\partial y}{\partial q}\frac{\partial y}{\partial l} \\ G &= \left(\frac{\partial x}{\partial l}\right)^2 + \left(\frac{\partial y}{\partial l}\right)^2 \end{aligned}\right\}, \tag{6.8}$$

we obtain:

$$\mathrm{d}s^2 = E(\mathrm{d}q)^2 + 2F(\mathrm{d}q)(\mathrm{d}l) + G(\mathrm{d}l)^2. \tag{6.9}$$

Substituting the above equation into (6.6) produces:

$$m^2 = \frac{E(\mathrm{d}q)^2 + 2F(\mathrm{d}q)(\mathrm{d}l) + G(\mathrm{d}l)^2}{r^2\left[(\mathrm{d}q)^2 + (\mathrm{d}l)^2\right]}. \tag{6.10}$$

Equation (6.10) does not include the direction-dependent elements. To introduce the condition that "scale factor m is independent of azimuth A", we need to further convert (6.10).

From Fig. 6.2, we have:

$$\tan(90^\circ - A) = \frac{P_1P_2}{PP_2} = \frac{M\mathrm{d}B}{r\mathrm{d}l} = \frac{\mathrm{d}q}{\mathrm{d}l},$$

and consequently:

$$\mathrm{d}l = \tan A\mathrm{d}q. \tag{6.11}$$

Inserting the above equation into (6.10) gives:

$$m^2 = \frac{E(\mathrm{d}q)^2 + 2F\tan A(\mathrm{d}q)^2 + G\tan^2 A(\mathrm{d}q)^2}{r^2\left[(\mathrm{d}q)^2 + \tan^2 A(\mathrm{d}q)^2\right]} = \frac{E + 2F\tan A + G\tan^2 A}{r^2\sec^2 A}$$
$$= \frac{E\cos^2 A + 2F\sin A\cos A + G\sin^2 A}{r^2}. \tag{6.12}$$

6.2.3 General Condition for Conformal Projection

To enable m to be independent of A in (6.12), we must satisfy the conditions that:

$$F = 0, E = \mathrm{G}.$$

Inserting (6.8) gives:

$$\left.\begin{aligned} &\frac{\partial x}{\partial q}\cdot\frac{\partial x}{\partial l} + \frac{\partial y}{\partial q}\cdot\frac{\partial y}{\partial l} = 0 \\ &\left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2 = \left(\frac{\partial x}{\partial l}\right)^2 + \left(\frac{\partial y}{\partial l}\right)^2 \end{aligned}\right\}. \tag{6.13}$$

From the first equation in (6.13), we have:

$$\frac{\partial x}{\partial l} = -\frac{\frac{\partial y}{\partial q}\frac{\partial y}{\partial l}}{\frac{\partial x}{\partial q}}.$$

Substituting into the second equation of (6.13) produces:

$$\left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2 = \frac{\left(\frac{\partial y}{\partial l}\right)^2}{\left(\frac{\partial x}{\partial q}\right)^2}\left[\left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2\right].$$

Removing the common term from both sides of the equation gives:

$$\left(\frac{\partial x}{\partial q}\right)^2 = \left(\frac{\partial y}{\partial l}\right)^2.$$

We evaluate the above equation and extract only the positive arguments to ensure that both the shape and the position are conformal after projection. Inserting into the first equation in (6.13) produces:

$$\left.\begin{aligned}\frac{\partial x}{\partial q} &= \frac{\partial y}{\partial l}\\ \frac{\partial x}{\partial l} &= -\frac{\partial y}{\partial q}\end{aligned}\right\}. \tag{6.14}$$

This is the general condition for the conformal projection from the ellipsoid onto the plane, which was derived by the French mathematician A.L. Cauchy and the German mathematician B. Riemann, and is therefore known as the Cauchy–Riemann differential equation.

The Cauchy–Riemann differential equation is the sufficient and necessary condition for a conformal projection. Likewise, the general condition for a conformal projection from the plane to the ellipsoid can be derived:

$$\left.\begin{aligned}\frac{\partial q}{\partial x} &= \frac{\partial l}{\partial y}\\ \frac{\partial l}{\partial x} &= -\frac{\partial q}{\partial y}\end{aligned}\right\}. \tag{6.15}$$

Also, after satisfying $F = 0$ and $E = G$, the formula for the scale factor (6.12) of a conformal projection from the ellipsoid onto the plane can be simplified to:

$$\left.\begin{aligned} m^2 &= \frac{E}{r^2} = \frac{\left(\frac{\partial x}{\partial q}\right)^2 + \left(\frac{\partial y}{\partial q}\right)^2}{r^2}, \\ \text{or} \qquad m^2 &= \frac{G}{r^2} = \frac{\left(\frac{\partial x}{\partial l}\right)^2 + \left(\frac{\partial y}{\partial l}\right)^2}{r^2}. \end{aligned}\right\} \tag{6.16}$$

These two formulae are equivalent, only of different form. We can choose either of these two formulae based on convenience for derivations. These formulae will be used to study the scale factor of the Gauss projection.

6.3 Fundamentals of the Gauss Projection

6.3.1 History and Development of the Gauss Projection

Gauss projection is the abbreviation for Gauss–Krüger projection, also known as the transverse cylindrical conformal projection, which is one of the conformal mappings of the Earth ellipsoid onto a plane. The Gauss projection was first credited to Carl Friedrich Gauss, a German mathematician, physicist, astronomer, and geodesist. During the period 1820–1830, when Gauss dealt with the triangulation results in Hannover, Germany, he used the ellipsoidal transverse Mercator projection that had a constant scale along the central meridian, which was a subject of his own investigations. However, he did not publish or release the results. The concluding formula for such a projection was found in the letters he wrote to his friends.

This projection was adapted by Oskar Schreiber in his book *Theory of Projection Methods of Hannover Land Surveying* (*Theorie der Projektionsmethode der hannoverschen Landsvermessung*) published in 1866 (Schreiber 1866) and thus brought the theory of the Gauss projection to the attention of the public.

A more detailed elaboration on the theory of Gauss projection and the practical formulae was provided by the German geodesist Louis Krüger in his *Conformal Mapping of the Earth Ellipsoid to the Plane* (*Konforme Abbildung des Erdellipsoids in der Ebene*) published in 1912 (Krüger 1912). Krüger thoroughly studied and complemented the Gauss projection in this book, which enabled such a projection to be widely used in many countries. This projection was therefore called the Gauss–Krüger projection, more usually known as the Gauss projection.

To apply the Gauss projection conveniently, the German scholar Boaga in 1919 suggested adoption of projection zones 3° wide. Each zone is assigned a false

y-coordinate of 500,000 m and the ordinates are also numbered. The division of projection zones starts from Greenwich.

The Gauss projection has received much attention and study from geodesists across the world, among which the investigations done by the Bulgarian geodesist Vladimir K. Khristov is the most representative. His two works *The Gauss-Krüger Coordinates on the Ellipsoid* (*Die Gauss-Krüger'schen Koordinaten auf dem Ellipsoid*) published in 1943 and *The Gaussian and Geographical Coordinates on the Krassowski Ellipsoid* (*Die Gaussschen und geographischen Koordinaten auf dem Ellipsoid von Krassowski*) published in 1955 enriched and developed Gauss projection in both theory and practice (Hristow 1943, 1955).

6.3.2 Conditions for Gauss Projection

In Fig. 6.3a we imagine an elliptical cylinder wrapped around the Earth ellipsoid tangential to a meridian on the ellipsoidal surface (which is called the central meridian or axial meridian). In addition, the central axis of the elliptic cylinder passes through the center of the ellipsoid. The ellipsoidal elements within a certain degree of longitude on either side of the central meridian are projected onto the elliptic cylindrical surface according to the three conditions given below. The cylindrical surface is then developed along the generating line passing through the north and south poles of the ellipsoid. The projection plane obtained is known as the Gauss projection plane, on which the central meridian and equator are projected as straight lines. The point of intersection of the projected central meridian and equator is taken as the coordinate origin O. The central meridian is labeled as the x-axis (north direction) of the projection, while the y-axis (east direction) is the mapping of the equator. Hence, the Gauss plane rectangular coordinate system is established; cf. Fig. 6.3b.

The three conditions for the Gauss projection are:

1. The projection is conformal.
2. The central meridian is projected as a straight line.
3. The length of the central meridian remains unchanged after projection.

The first condition is the general condition for conformal projection, while the latter two are the particular conditions of the Gauss projection itself. We use mathematical relations to express the three conditions:

1. $\dfrac{\partial x}{\partial q} = \dfrac{\partial y}{\partial l}, \dfrac{\partial x}{\partial l} = -\dfrac{\partial y}{\partial q}$;
2. When $l = 0$, $y = 0$
3. When $l = 0$, $x = X$ (X denotes the length of the meridian arc from the equator)

In terms of solving differential equations, condition 1 can only find the general solutions of the equation. To find the particular solutions of the differential equation, one also needs to plug in the initial conditions, i.e. conditions 2 and 3. How to

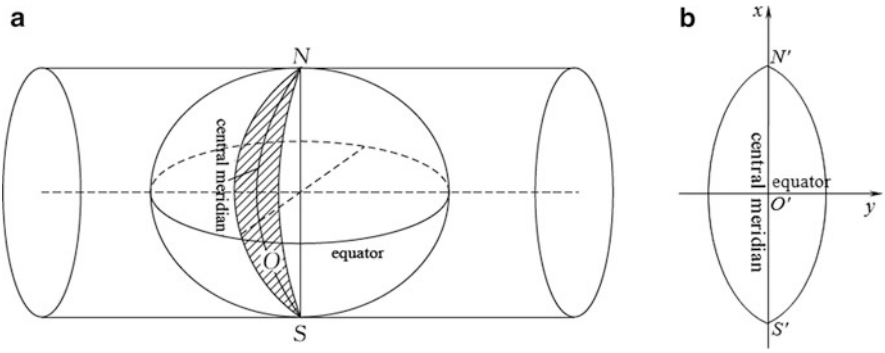


Fig. 6.3 Geometric description of the Gauss projection (**a**) elliptical cylinder being tangential to the ellipsoid and (**b**) developed plane of the elliptic cylinder

determine the projection formula (6.1) according to these three conditions for Gauss projection will be discussed in Sect. 6.4.

6.3.3 Zone-Dividing of the Gauss Projection

Reasons for Zone Division

In the Gauss projection, except for the central meridian, any other line segments will be distorted after projection and the distortion increases with the distance from the central meridian. Such distortion of distance is harmful and should be kept within small limits to reduce its effect on mapping and map application. The most effective way to limit the distortion of distance is by "zone-dividing" projection. To be specific, the globe is divided into narrow zones along meridians that are all of equal degrees width in longitude (e.g., 6° or 3° wide in longitude). Each zone is projected separately to get several different projection zones. The meridian at the center of each zone is called the central meridian, which is projected as the ordinate axis. Hence, the central meridian is also known as the axial meridian. The meridian used to divide the projection zones (at the edge of a projection zone) is called the dividing meridian.

Because a zone division limits the projection area to within a narrow area of the two sides of the central meridian, the distance distortion is effectively restricted. Evidently, within a certain area, zones become narrower as the number of zones increases, and the distortion of distances will correspondingly become smaller. In terms of limiting distortions of distance, the more zones divided, the better.

After the zone-dividing projection, each projection zone has unique coordinate axes and coordinate system origins. Thus, a mutually independent Gauss plane coordinate system has formed and points on the two sides of the dividing meridian belong to two different coordinate systems. In productive operations, the areas of

interest usually cross between different zones and have to be transformed into the same coordinate system, so a transformation of coordinates between different projection zones is necessary (called the transformation between adjacent zones). Considering this, the area of interest should not be divided into too many zones.

When dividing an area into zones, both of the above-mentioned aspects should be taken into account. There are two methods of zone division; namely, that of the zones 6° wide (each zone is 6° of longitude in width) and that of the zones 3° wide (each zone is 3° of longitude in width). The former can be used for medium- and small-scale mapping and the latter for large-scale mapping. The plane rectangular coordinates of the geodetic points should be computed within the zones 6° wide according to the Gauss projection. In areas with 1:10,000 or much larger scale mapping, the plane rectangular coordinates in the 3° zones should also be computed.

Methods of Zone Division

Figure 6.4 depicts the Gauss projection in zones 6° wide. Starting from 0° meridian eastward, each zone is 6° of longitude in width, numbered 1 to 60. The central meridians in each zone are of longitudes 3°, 9°, . . ., up to 357°. We assume that the zone is numbered n, and the central meridian is of longitude L_0, hence:

$$\left.\begin{aligned} L_0 &= n \cdot 6^\circ - 3^\circ \\ n &= \frac{(L_0 + 3^\circ)}{6} \end{aligned}\right\}. \tag{6.17}$$

Given the geodetic longitude L of a certain point, the zone number of this point for a projection in zones of 6° width can be computed according to:

$$n = \frac{L}{6}(\text{take integer quotient}) + 1\ (\text{if there is a remainder}).$$

The 3° zones are divided based on the 6° zones. The central meridians in even-numbered zones coincide with those in the zones 6° wide. The central meridians in odd-numbered zones coincide with the zone-dividing meridians in the zones 6° wide. The specific zone dividing starts from the meridian of east longitude 1.5° eastward; each zone is 3° of longitude in width, numbered 1 to 120 as shown in Fig. 6.4. Setting the zone number to n', the longitudes of the central meridians in each zone are:

$$\left.\begin{aligned} L_0 &= 3^\circ n' \\ n' &= L_0/3 \end{aligned}\right\} \tag{6.18}$$

Given the geodetic longitude L of a point, the zone number of this point for a projection in zones of 3° width can be computed according to:

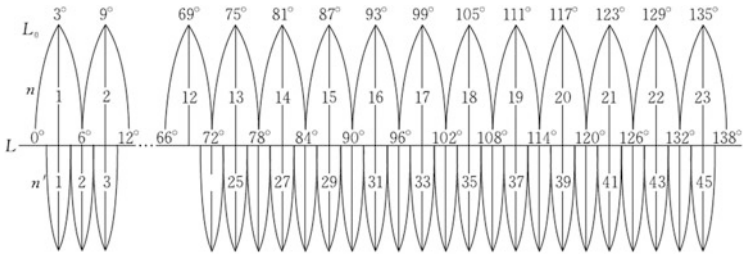


Fig. 6.4 Zone division of the Gauss projection

$$n' = \frac{L - 1.5}{3}(\text{take integer quotient}) + 1.$$

Overlap Between Projection Zones

After a zone-dividing projection, the rectangular coordinate systems in the adjacent zones are mutually independent. To carry out control network adjustment across zones, to plot or use topographic maps across the zone boundaries, to show the control points outside the mapping area (situated in the neighboring projection zones), etc., the neighboring projection zones have to be overlapped to a certain degree, as shown in Fig. 6.5.

The so-called overlap between projection zones means that the control points within a certain area have the coordinate values of two adjacent zones. Thus, on the topographic maps in this area there are two sets of kilometer grids (the grid of the coordinate system in its own zone and that in the adjacent zone).

For the overlap of projection zones, China has the following regulations: A 30-min longitude overlap is allowed between zones from west to east (equivalent to the longitude width of a 1:100,000 map); and a 15-min longitude overlap is allowed between zones from east to west (corresponding to the longitude width of a 1:50,000 map). Namely, each projection zone extends 30′ to the east and 15′ to the west; thus, an overlap of 45′ in longitude is formed near the dividing meridian, as shown in Fig. 6.5.

6.3.4 Natural Coordinates and False (Biased) Coordinates

In the northern hemisphere, after zone-dividing projection, all x-values of Gauss coordinates are positive and y can have both positive and negative values. To avoid a negative value of y, 500,000 m is assigned to it, which corresponds to a false or biased y-coordinate of 500,000 m. Hence, the coordinate values of y are all positive. Also, each zone is projected separately and forms a mutually independent plane rectangular coordinate system. The same coordinate value (x, y) in each zone will have a point corresponding to it, which is likely to cause confusion about the

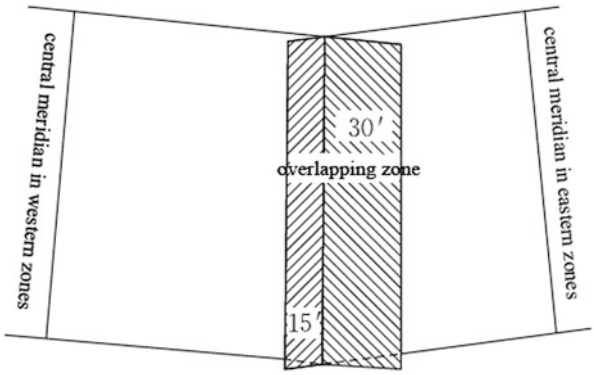


Fig. 6.5 Overlap between projection zones

position of a point. To describe the zones in which a point lies, a zone number is designated before the y value (assigned 500,000 m). The coordinates formed in such way are called false coordinates, denoted by y_{false}. To show the results, the points are all expressed by the false coordinates, whereas in real applications, the zone number should be eliminated and the 500,000 m value subtracted. The recovered values are called the natural coordinates of this point. The relationship between natural coordinates and false coordinates is shown in Fig. 6.6.

For instance, in 6° zone 19, the natural coordinates of points A and B are:

$$A: \begin{cases} x = 4\ 485\ 076.81\ \text{m} \\ y = -2\ 578.86\ \text{m} \end{cases} \quad B: \begin{cases} x = 4\ 485\ 076.81\ \text{m} \\ y = 2\ 578.86\ \text{m} \end{cases}$$

and their false coordinates are:

$$A: \begin{cases} x = 4\ 485\ 076.81\ \text{m} \\ y_{\text{false}} = 19\ 497\ 421.14\ \text{m} \end{cases} \quad B: \begin{cases} x = 4\ 485\ 076.81\ \text{m} \\ y_{\text{false}} = 19\ 502\ 578.86\ \text{m} \end{cases}$$

6.4 Direct and Inverse Solutions of the Gauss Projection and Transformation Between Adjacent Zones

6.4.1 Formula for Direct Solution of the Gauss Projection

Derivations of the Formula

The formula for the direct solution of the Gauss projection is the formula used to compute the Gauss plane rectangular coordinates (x, y), given the geodetic coordinates (L, B) or (l, q) of a point, as shown in Fig. 6.7.

The projection equation from the ellipsoid to the plane is generally expressed as:

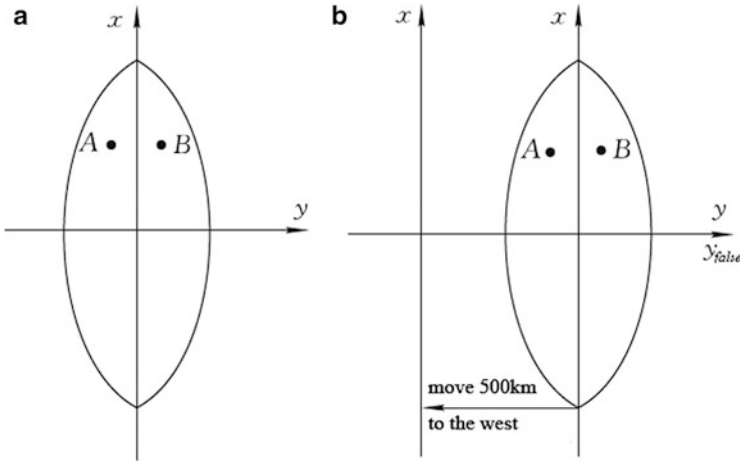


Fig. 6.6 (**a**) Natural coordinates and (**b**) false coordinates (biased by 500,000 m)

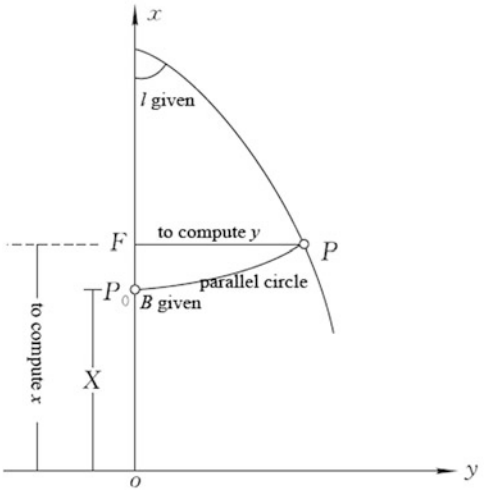


Fig. 6.7 Direct solution of the Gauss projection

$$\left.\begin{aligned} x &= f_1(l,q) \\ y &= f_2(l,q) \end{aligned}\right\}. \tag{6.19}$$

The Gauss projection has three conditions, from which we can derive the formula for a direct solution of the Gauss projection.

A Gauss projection is carried out along the narrow strips within a certain longitude east and west of the central meridian. In each projection area, the difference l between the longitude of an arbitrary point and that of the central

meridian is quite small, generally within 0–3.5°. Besides, the arc value $\frac{l}{\rho}$ is a small quantity, so we can expand (6.19) at point (0, q) according to a Taylor series expansion of binary functions. In this case, the increments of the independent variable of an arbitrary point (l, q) relative to the expansion point (0, q) are l and 0, respectively, i.e., the increment of q relative to the expansion point is 0. Hence, the partial derivative with respect to q and the term that includes mixed partial derivatives with respect to q are all 0. The x series is expanded as:

$$\begin{aligned} x &= f_1(0,q) + \left.\frac{\partial f_1}{\partial l}\right|_{(0,q)} l + \left.\frac{1}{2!}\frac{\partial^2 f_1}{\partial l^2}\right|_{(0,q)} l^2 + \left.\frac{1}{3!}\frac{\partial^3 f_1}{\partial l^3}\right|_{(0,q)} l^3 + \left.\frac{1}{4!}\frac{\partial^4 f_1}{\partial l^4}\right|_{(0,q)} l^4 + \ldots \\ &= m_0 + m_1 l + m_2 l^2 + m_3 l^3 + m_4 l^4 + \ldots, \end{aligned}$$

which is the power series of the difference in longitude l. The value of each partial derivative at point (0, q) no longer includes the variable l. Hence $m_i (i = 0, 1, 2, 3, 4,....)$ is the function of isometric latitude q; $m_0 = f_1(0, q)$ indicates the x coordinate of the point (0, q) on the central meridian. From the third condition for the Gauss projection (when $l = 0$, $x = X$), one can get $m_0 = X$.

By the same token, we have:

$$\begin{aligned} y &= f_2(0,q) + \left.\frac{\partial f_2}{\partial l}\right|_{(0,q)} l + \left.\frac{1}{2!}\frac{\partial^2 f_2}{\partial l^2}\right|_{(0,q)} l^2 + \left.\frac{1}{3!}\frac{\partial^3 f_2}{\partial l^3}\right|_{(0,q)} l^3 + \left.\frac{1}{4!}\frac{\partial^4 f_2}{\partial l^4}\right|_{(0,q)} l^4 + \ldots \\ &= n_0 + n_1 l + n_2 l^2 + n_3 l^3 + n_4 l^4 + \ldots, \end{aligned}$$

which is the power series of the difference in longitude l. $n_i (i = 0, 1, 2, 3, 4,....)$ is the function of isometric latitude q; $n_0 = f_2(0, q)$ indicates the y coordinate of the point (0, q) on the central meridian. According to the second condition for Gauss projection (when $l = 0$, $y = 0$), we have $n_0 = 0$.

Equation (6.19) has been expanded into the power series of the longitude difference l:

$$\left.\begin{aligned} x &= X + m_1 l + m_2 l^2 + m_3 l^3 + m_4 l^4 + \ldots \\ y &= n_1 l + n_2 l^2 + n_3 l^3 + n_4 l^4 + \ldots \end{aligned}\right\}, \tag{6.20}$$

where $m_1, m_2, \ldots, n_1, n_2, \ldots$ are undetermined coefficients. They are the functions of the isometric latitude q (or geodetic latitude B). To apply the first condition for Gauss projection, taking the partial derivative of (6.20) yields:

$$\left.\begin{aligned}
\frac{\partial x}{\partial q} &= \frac{\mathrm{d}X}{\mathrm{d}q} + l\frac{\mathrm{d}m_1}{\mathrm{d}q} + l^2\frac{\mathrm{d}m_2}{\mathrm{d}q} + l^3\frac{\mathrm{d}m_3}{\mathrm{d}q} + l^4\frac{\mathrm{d}m_4}{\mathrm{d}q} + \ldots \\
\frac{\partial x}{\partial l} &= m_1 + 2m_2 l + 3m_3 l^2 + 4m_4 l^3 + \ldots \\
\frac{\partial y}{\partial q} &= l\frac{\mathrm{d}n_1}{\mathrm{d}q} + l^2\frac{\mathrm{d}n_2}{\mathrm{d}q} + l^3\frac{\mathrm{d}n_3}{\mathrm{d}q} + l^4\frac{\mathrm{d}n_4}{\mathrm{d}q} + \ldots \\
\frac{\partial y}{\partial l} &= n_1 + 2n_2 l + 3n_3 l^2 + 4n_4 l^3 + \ldots
\end{aligned}\right\}. \tag{6.21}$$

Inserting the first condition for Gauss projection, namely the general condition for conformal projection $\frac{\partial x}{\partial q} = \frac{\partial y}{\partial l}$ and $\frac{\partial x}{\partial l} = -\frac{\partial y}{\partial q}$, gives:

$$\frac{\mathrm{d}X}{\mathrm{d}q} + l\frac{\mathrm{d}m_1}{\mathrm{d}q} + l^2\frac{\mathrm{d}m_2}{\mathrm{d}q} + l^3\frac{\mathrm{d}m_3}{\mathrm{d}q} + l^4\frac{\mathrm{d}m_4}{\mathrm{d}q} + \ldots = n_1 + 2n_2 l + 3n_3 l^2 + 4n_4 l^3 + \ldots,$$

$$m_1 + 2m_2 l + 3m_3 l^2 + 4m_4 l^3 + \ldots = -l\frac{\mathrm{d}n_1}{\mathrm{d}q} - l^2\frac{\mathrm{d}n_2}{\mathrm{d}q} - l^3\frac{\mathrm{d}n_3}{\mathrm{d}q} - l^4\frac{\mathrm{d}n_4}{\mathrm{d}q} + \ldots$$

To make the right sides of the above two equations equal, the necessary and sufficient condition is that the coefficients of the same powers of l are equal. Hence:

$$\left.\begin{aligned}
&m_1 = n_2 = m_3 = n_4 = \cdots = 0 \\
&n_1 = \frac{\mathrm{d}X}{\mathrm{d}q} \\
&m_2 = -\frac{1}{2}\frac{\mathrm{d}n_1}{\mathrm{d}q} \\
&n_3 = \frac{1}{3}\frac{\mathrm{d}m_2}{\mathrm{d}q} \\
&m_4 = -\frac{1}{4}\frac{\mathrm{d}n_3}{\mathrm{d}q} \\
&n_5 = \frac{1}{5}\frac{\mathrm{d}m_4}{\mathrm{d}q} \\
&\vdots
\end{aligned}\right\}. \tag{6.22}$$

Since $n_0 = m_1 = n_2 = m_3 = n_4 = \cdots = 0$, (6.20) can be simplified as:

$$\left.\begin{aligned}
x &= X + m_2 l^2 + m_4 l^4 + \ldots \\
y &= n_1 l + n_3 l^3 + n_5 l^5 + \ldots
\end{aligned}\right\}. \tag{6.23}$$

It can be seen from the above equation that the projections on the east and west sides of the central meridian are symmetrical about the central meridian.

To determine the coefficients of $n_1, m_2, n_3, m_4, n_5, \ldots$ it is necessary to find the derivative $\frac{\mathrm{d}X}{\mathrm{d}q}$. It follows from the differential formula for meridian arc length $\mathrm{d}X = M\mathrm{d}B$ and (6.3) $\frac{\mathrm{d}B}{\mathrm{d}q} = \frac{r}{M}$ that:

$$n_1 = \frac{\mathrm{d}X}{\mathrm{d}q} = \frac{\mathrm{d}X}{\mathrm{d}B}\frac{\mathrm{d}B}{\mathrm{d}q} = r.$$

Hence, it follows that:

$$n_1 = r = N\cos B = \frac{c}{V}\cos B. \tag{6.24}$$

We obtain $\frac{\mathrm{d}n_1}{\mathrm{d}q}$ by:

$$\frac{\mathrm{d}n_1}{\mathrm{d}q} = \frac{\mathrm{d}r}{\mathrm{d}q} = \frac{\mathrm{d}r}{\mathrm{d}B}\frac{\mathrm{d}B}{\mathrm{d}q} = \frac{\mathrm{d}}{\mathrm{d}B}\left(\frac{c}{V}\cos B\right)\frac{\mathrm{d}B}{\mathrm{d}q} = \left(-\frac{c}{V^2}\frac{\mathrm{d}V}{\mathrm{d}B}\cos B - \frac{c}{V}\sin B\right)\frac{\mathrm{d}B}{\mathrm{d}q},$$

with

$$\frac{\mathrm{d}V}{\mathrm{d}B} = \frac{\mathrm{d}}{\mathrm{d}B}\left(1 + e'^2\cos^2 B\right)^{\frac{1}{2}} = \frac{1}{2}\left(1 + e'^2\cos^2 B\right)^{-\frac{1}{2}}\frac{\mathrm{d}}{\mathrm{d}B}\left(e'^2\cos^2 B\right)$$

$$= \frac{1}{V}e'^2\cos B(-\sin B) = -\frac{1}{V}e'^2\cos^2 B\frac{\sin B}{\cos B}.$$

Some of the formulae in this chapter are quite long, so to simplify writing and improve readability, we have intentionally introduce some symbols as follows:

$$\left.\begin{array}{l}\eta = e'\cos B\\ t = \tan B\end{array}\right\}.$$

Hence, with:

$$\frac{\mathrm{d}V}{\mathrm{d}B} = -\frac{1}{V}\eta^2 t,$$

we get:

$$\frac{\mathrm{d}n_1}{\mathrm{d}q} = \left[-\frac{c}{V^2}\left(-\frac{1}{V}\eta^2 t \right) \cos B - \frac{c}{V}\sin B \right] \frac{\frac{c}{V}\cos B}{\frac{c}{V^3}}$$

$$= \left[\frac{c}{V^3}\sin B\left(\eta^2 - V^2\right) \right] V^2 \cos B = \left[-\frac{c}{V^3}\sin B \right] V^2 \cos B$$

$$= -\frac{c}{V}\sin B\cos B = -N\sin B\cos B.$$

Inserting into the third equation in (6.22) gives:

$$m_2 = \frac{N}{2}\sin B\cos B. \tag{6.25}$$

Taking derivatives in turn from m_2, and substituting correspondingly into (6.22), we get n_3, m_4, n_5 ... as:

$$\left.\begin{aligned} n_3 &= \frac{N}{6}\cos^3 B\left(1 - t^2 + \eta^2\right) \\ m_4 &= \frac{N}{24}\sin B\cos^3 B\left(5 - t^2 + 9\eta^2\right) \\ n_5 &= \frac{N}{120}\cos^5 B\left(5 - 18t^2 + t^4\right) \\ &\vdots \end{aligned}\right\}. \tag{6.26}$$

Substituting (6.24), (6.25), and (6.26) into (6.23) and neglecting terms like $\eta^2 l^5$ and l^6 or higher orders, produces the formula for the direct solution of the Gauss projection:

$$\left.\begin{aligned} x &= X + \frac{N}{2\rho''^2}\sin B\cos B l''^2 + \frac{N}{24\rho''^4}\sin B\cos^3 B\left(5 - t^2 + 9\eta^2\right) l''^4 \\ y &= \frac{N}{\rho''}\cos B l'' + \frac{N}{6\rho''^3}\cos^3 B\left(1 - t^2 + \eta^2\right) l''^3 + \frac{N}{120\rho''^5}\cos^5 B\left(5 - 18t^2 + t^4\right) l''^5 \end{aligned}\right\}, \tag{6.27}$$

where $\rho'' = 206264.80625''$. l'' (unit: second) is the difference between the longitude of point P on the ellipsoid and that of the central meridian. l is positive when point P is east of the central meridian, and negative when P is west of the central meridian. B is the geodetic latitude of P. X is the length of the meridian arc from the equator to point of latitude B. Given the geodetic coordinates (L, B) of point P (the

longitude of the central meridian L_0 is given, and $l = L - L_0$ can be computed), the Gauss plane coordinates (x, y) of P can be calculated according to (6.27).

The functional relationship between (x, y) and (L, B) expressed in (6.27) has determined the specific form of f_1 and f_2 in (6.19).

When $l < 3.5^\circ$, the computation according to (6.34) is accurate to 0.1 m. If the desired accuracy is 0.001 m, the series in (6.27) can be further expanded. The process is not shown here. The computational formula is directly given as:

$$\left.\begin{aligned} x &= X + \frac{N}{2\rho''^2}\sin B\cos B\cdot l''^2 + \frac{N}{24\rho''^4}\sin B\cos^3 B\left(5 - t^2 + 9\eta^2 + 4\eta^4\right)l''^4 \\ &\quad + \frac{N}{720\rho''^6}\sin B\cos^5 B\left(61 - 58t^2 + t^4\right)l''6 \\ y &= \frac{N}{\rho''}\cos Bl'' + \frac{N}{6\rho''^3}\cos^3 B\left(1 - t^2 + \eta^2\right)l''^3 + \frac{N}{120\rho''^5}\cos^5 B\left(5 - 18t^2\right. \\ &\quad \left. + t^4 + 14\eta^2 - 58\eta^2 t^2\right)l''^5 \end{aligned}\right\}. \tag{6.28}$$

For the most accurate formulae and for new methods, some literature is recommended for further reading (Karney 2011; Kawase 2011; Kawase 2012; Deakin, et al. 2011).

Formula Analysis

Analyzing (6.27), one can get the shapes of the meridians and parallels on the ellipsoid after the projection (cf. Fig. 6.8).

Projections of the Central Meridian and Equator

When $B = 0$, $x = 0$, and y changes with l, it indicates that the equator is projected as a straight line, i.e., the abscissa axis. When $l = 0$, $y = 0$, $x = X$ it indicates that the central meridian is also projected as a straight line, i.e., the ordinate axis. There are no distortions in the projection. The point of intersection of the projected central meridian and equator is the origin of the plane coordinate system.

Meridian Projection

Setting l as constant, we can get the parameter equation of curves of the projected meridians in terms of parameter B. When the value of B increases, x increases while

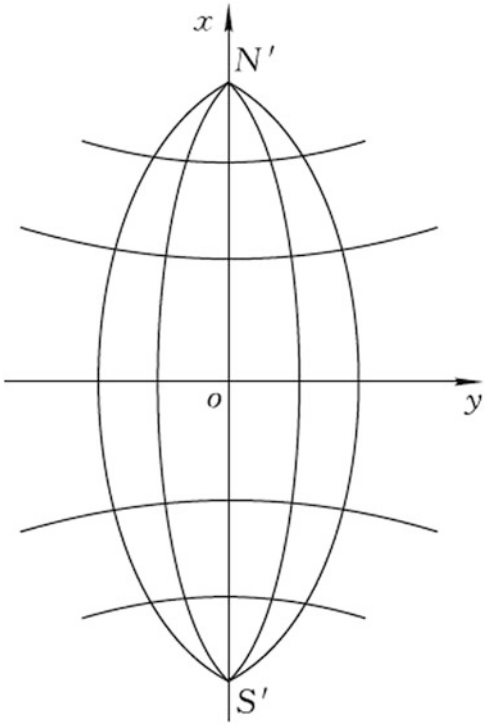


Fig. 6.8 Shapes of the meridian and parallel projections

y decreases; when B is negative, $\sin(-B) = -\sin B$ and $\cos(-B) = \cos B$; thus the opposite sign of value x equals the value of y. Hence, the projected meridians curve towards the central meridian and converge towards the poles. The meridian projections are also symmetrical with respect to the central meridian and equator.

Parallel Projection

Here, $B =$ constant and x and y vary only with l. When the value of l increases, that of x and y also increases; when l is negative, the value of x is the same while the value of y has an opposite sign because x is the even power function of l and y is the odd power function of l. So, the parallels are projected as curves that are symmetrical in relation to the x-axis and bending towards the poles.

Symmetry of the Meridian and Parallel Projections

To sum up, the meridians and parallels on the ellipsoid symmetrical about the central meridian and equator are projected as curves symmetrical with respect to the x-axis and y-axis.

Projection of an Arbitrary Geodesic

Arbitrary geodesics are projected as curves that are concave towards the central meridian and poles.

Distortion of Projection

It can be seen from Fig. 6.8 that the projected meridians appear more curved and distorted the further away they are from the central meridian. Thus, distortion increases further away from the central meridian.

Practical Formulae

This section provides the practical formula for the direct solution of the Gauss Projection, which is suitable for computations based on computer programming. Parameters relevant to the Krassowski Ellipsoid, GRS75 Ellipsoid, and GRS80 Ellipsoid, respectively, are also provided for practical use.

Formulae for the Direct Solution of the Gauss Projection (Accurate to 0.001 m)

In (6.28), we set $m = \cos B \cdot l^{\circ} \cdot \dfrac{\pi}{180^{\circ}}$ to obtain:

$$\left.\begin{aligned} x &= X + Nt\left[\frac{1}{2}m^2 + \frac{1}{24}\left(5 - t^2 + 9\eta^2 + 4\eta^4\right)m^4 + \frac{1}{720}\left(61 - 58t^2 + t^4\right)m^6\right] \\ y &= N\left[m + \frac{1}{6}\left(1 - t^2 + \eta^2\right)m^3 + \frac{1}{120}\left(5 - 18t^2 + t^4 + 14\eta^2 - 58\eta^2 t^2\right)m^5\right] \end{aligned}\right\}, \tag{6.29}$$

where the meridian arc with length X is computed according to (5.41) if the Krassowski Ellipsoid is adopted; if the GRS75 or GRS80 Ellipsoid is adopted, X is computed according to (5.42) or (5.43), respectively. Computations continue to the eighth-power term in (5.41), (5.42), or (5.43).

Formulae for the Direct Solution of the Gauss Projection (Accurate to 0.1 m)

In (6.27), we set $m = \cos B \cdot l^{\circ} \cdot \dfrac{\pi}{180^{\circ}}$ to obtain:

$$\left.\begin{aligned} x &= X + Nt\left[\frac{1}{2}m^2 + \frac{1}{24}\left(5 - t^2 + 9\eta^2 + 4\eta^4\right)m^4\right] \\ y &= N\left[m + \frac{1}{6}\left(1 - t^2 + \eta^2\right)m^3 + \frac{1}{120}\left(5 - 18t^2 + t^4\right)m^5\right] \end{aligned}\right\}, \tag{6.30}$$

where the meridian arc with length X is computed also according to (5.41), (5.42), and (5.43), respectively, according to different ellipsoidal parameters. Computations continuing to the sixth-power term will be sufficient.

Example

Relevant parameters used in the practical formulae for the Krassowski, Ellipsoid, GRS75 Ellipsoid, and GRS80 Ellipsoid are given in Sect. 5.2.2.

We take the computer programmed computation of (6.29) as an example; cf. Table 6.1.

6.4.2 *Formula for Inverse Solution of the Gauss Projection*

Derivations of the Formula

As Fig. 6.9 shows, the formulae for the inverse solution of the Gauss projection are those used to compute the geodetic coordinates (L, B) or the corresponding (l, q) with given Gauss plane coordinates (x, y) of point P.

The equation of projection from plane to ellipsoid is:

$$\left.\begin{aligned} q &= f'_1(x, y) \\ l &= f'_2(x, y) \end{aligned}\right\}. \tag{6.31}$$

Analogous to the derivations of the formulae for the direct solution of the Gauss projection, we expand the equation of projection (6.31) from the plane onto the ellipsoid into the power series and determine the specific forms of the projection function f'_1 and f'_2 using the method of undetermined coefficients according to the three conditions for Gauss projection. Therefore, the formulae for the inverse solution of the Gauss projection will be derived.

The value of y at point P is small compared to the radius of the ellipsoid. So, the function in (6.31) can be expanded into the power series of y. The expansion point is the point F (x, 0), which is the foot of the perpendicular from point P to the central meridian, also known as the foot point. The latitude of this point is called the

Table 6.1 Direct solution of the Gauss projection: sample computation

Given data	Ellipsoidal parameters	Computational results (m)	
		6° zones	3° zones
$B = 40° 58' 32.33''$	Krassowski Ellipsoid	$x = 4{,}538{,}610.951$	$x = 4{,}538{,}610.951$
$L = 100° 10' 20.11''$		$y = 98{,}666.625$	$y = 98{,}666.625$
		$y_{\text{false}} = 17, 598, 666.625$	$y_{\text{false}} = 33, 598, 666.625$
	GRS75 Ellipsoid	$x = 4{,}538{,}532.847$	$x = 4{,}538{,}532.847$
		$y = 98{,}665.022$	$y = 98{,}665.022$
		$y_{\text{false}} = 17, 598, 665.022$	$y_{\text{false}} = 33, 598, 665.022$
	GRS80 Ellipsoid	$x = 4{,}538{,}530.729$	$x = 4{,}538{,}530.729$
		$y = 98{,}664.975$	$y = 98{,}664.975$
		$y_{\text{false}} = 17\ 598\ 664.975$	$y_{\text{false}} = 33, 598\ 664.975$
$B = 35° 26' 40.38''$	Krassowski Ellipsoid	$x = 3{,}925{,}560.035$	$x = 3{,}924{,}588.054$
$L = 115° 08' 51.22''$		$y = -168{,}198.578$	$y = 104{,}193.075$
		$y_{\text{false}} = 20, 331, 801.422$	$y_{\text{false}} = 38, 604, 193.075$
	GRS75 Ellipsoid	$x = 3{,}925{,}492.277$	$x = 3{,}924{,}520.313$
		$y = -168{,}195.836$	$y = 104{,}191.377$
		$y_{\text{false}} = 20, 331, 804.164$	$y_{\text{false}} = 38, 604, 191.377$
	GRS80 Ellipsoid	$x = 3{,}925{,}490.447$	$x = 3{,}924{,}518.483$
		$y = -168{,}195.757$	$y = 104{,}191.328$
		$y_{\text{false}} = 20, 331, 804.243$	$y_{\text{false}} = 38, 604, 191.328$

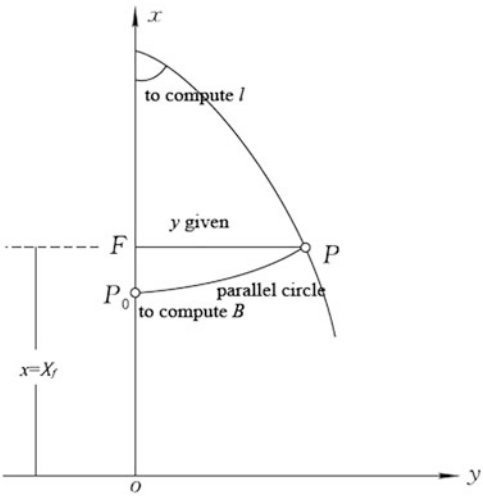


Fig. 6.9 Inverse solution of Gauss projection

footprint latitude and is denoted by B_f, whose corresponding isometric latitude is q_f. The meridian arc length from the equator to B_f is X_f; namely, the ordinate of point F is $x = X_f$, and the value of B_f can be obtained reversely from X_f according to the formula for meridian arc length.

According to the symmetry property of the Gauss projection and the second condition for the projection, the series can be directly written according to (6.23):

$$\left.\begin{aligned} q &= m'_0 + m'_2 y^2 + m'_4 y^4 + \cdots \\ l &= n'_1 y + n'_3 y^3 + n'_5 y^5 + \cdots \end{aligned}\right\}. \tag{6.32}$$

Taking the partial derivative of (6.32) based on the first condition for the Gauss projection yields:

$$\left.\begin{aligned} \frac{\partial q}{\partial x} &= \frac{\mathrm{d}m'_0}{\mathrm{d}x} + y^2 \frac{\mathrm{d}m'_2}{\mathrm{d}x} + y^4 \frac{\mathrm{d}m'_4}{\mathrm{d}x} + \cdots \\ \frac{\partial q}{\partial y} &= 2m'_2 y + 4m'_4 y^3 + \cdots \\ \frac{\partial l}{\partial x} &= y \frac{\mathrm{d}n'_1}{\mathrm{d}x} + y^3 \frac{\mathrm{d}n'_3}{\mathrm{d}x} + y^5 \frac{\mathrm{d}n'_5}{\mathrm{d}x} + \cdots \\ \frac{\partial l}{\partial y} &= n'_1 + 3n'_3 y^2 + 5n'_5 y^4 + \cdots \end{aligned}\right\}. \tag{6.33}$$

Inserting the general condition for conformal projection gives

$$\begin{aligned} \frac{\mathrm{d}m'_0}{\mathrm{d}x} + y^2 \frac{\mathrm{d}m'_2}{\mathrm{d}x} + y^4 \frac{\mathrm{d}m'_4}{\mathrm{d}x} + \cdots &= n'_1 + 3n'_3 y^2 + 5n'_5 y^4 + \cdots \\ y \frac{\mathrm{d}n'_1}{\mathrm{d}x} + y^3 \frac{\mathrm{d}n'_3}{\mathrm{d}x} + y^5 \frac{\mathrm{d}n'_5}{\mathrm{d}x} + \cdots &= -2m'_2 y - 4m'_4 y^3 - \cdots \end{aligned}$$

The above equation must satisfy the condition that the coefficients of the same powers of y are equal in order to be valid; hence:

$$\left.\begin{aligned} n'_1 &= \frac{\mathrm{d}m'_0}{\mathrm{d}x} \\ m'_2 &= -\frac{1}{2}\frac{\mathrm{d}n'_1}{\mathrm{d}x} \\ n'_3 &= \frac{1}{3}\frac{\mathrm{d}m'_2}{\mathrm{d}x} \\ m'_4 &= -\frac{1}{4}\frac{\mathrm{d}n'_3}{\mathrm{d}x} \\ n'_5 &= \frac{1}{5}\frac{\mathrm{d}m'_4}{\mathrm{d}x} \\ &\vdots \end{aligned}\right\}. \tag{6.34}$$

To obtain the above derivative, m'_0 needs to be determined first. According to the third condition for the Gauss projection, with $y = 0$, $x = X_f$. In such a case, we

assume that the isometric latitude corresponding to X_f is q_f, and it follows from the first equation in (6.32) that:

$$q = m'_0 = q_f \tag{6.35}$$

Inserting into (6.34) results in:

$$\begin{aligned} n'_1 &= \frac{\mathrm{d}q_f}{\mathrm{d}x} = \left(\frac{\mathrm{d}q}{\mathrm{d}x}\right)_f = \left(\frac{\mathrm{d}q}{\mathrm{d}B}\frac{\mathrm{d}B}{\mathrm{d}X}\right)_f = \left(\frac{M}{N\cos B}\frac{1}{M}\right)_f = \frac{1}{N_f \cos B_f} \\ &= \frac{\sec B_f}{N_f}, \end{aligned} \tag{6.36}$$

where x is written as X, which is valid only when $y = 0$, i.e., $q = q_f$. Hence, while taking derivatives, when x is alternatively written as X, the subscript f will be used outside the brackets of the derivatives.

Taking derivatives in turn and inserting into (6.34) produces m'_2, n'_3, m'_4, and m'_5, etc.:

$$\left.\begin{aligned} m'_2 &= -\frac{t_f \sec B_f}{2{N_f}^2} \\ n'_3 &= -\frac{\sec B_f}{6{N_f}^3}\left(1 + 2{t_f}^2 + {\eta_f}^2\right) \\ m'_4 &= \frac{t_f \sec B_f}{24{N_f}^4}\left(5 + 6{t_f}^2 + {\eta_f}^2\right) \\ n'_5 &= \frac{\sec B_f}{120{N_f}^5}\left(5 + 28{t_f}^2 + 24{t_f}^4\right) \end{aligned}\right\}. \tag{6.37}$$

Substituting (6.35), (6.36), and (6.37) into (6.32) yields:

$$\left.\begin{aligned} q &= q_f - \frac{t_f \sec B_f}{2{N_f}^2}y^2 + \frac{t_f \sec B_f}{24{N_f}^4}\left(5 + 6{t_f}^2 + {\eta_f}^2\right)y^4 \\ l &= \frac{\sec B_f}{N_f}y - \frac{\sec B_f}{6{N_f}^3}\left(1 + 2{t_f}^2 + {\eta_f}^2\right)y^3 + \frac{\sec B_f}{120{N_f}^5}\left(5 + 28{t_f}^2 + 24{t_f}^4\right)y^5 \end{aligned}\right\}. \tag{6.38}$$

The above steps and methods for derivations are the same as those for the direct solution. Nevertheless, at this moment, we only get the isometric latitude q. To obtain the geodetic latitude B, we need to further convert the first equation in (6.38).

As mentioned before, the geodetic latitude and isometric latitude are related. Assume that their functional relationship is:

$$B = F(q). \tag{6.39}$$

Similarly, we have:

$$B_f = F(q_f). \tag{6.40}$$

We expand $B = F(q) = F\left(q_f + \overline{q - q_f}\right)$, and according to the Taylor series we obtain:

$$B = F(q_f) + \left(\frac{\mathrm{d}B}{\mathrm{d}q}\right)_f (q - q_f) + \frac{1}{2!}\left(\frac{\mathrm{d}^2 B}{\mathrm{d}q^2}\right)_f (q - q_f)^2 + \ldots \tag{6.41}$$

It follows from (6.3) that:

$$\left.\begin{aligned} \left(\frac{\mathrm{d}B}{\mathrm{d}q}\right)_f &= \left(\frac{r}{M}\right)_f = \left(\frac{N\cos B}{M}\right)_f = {V_f}^2 \cos B_f \\ \left(\frac{\mathrm{d}^2 B}{\mathrm{d}q^2}\right)_f &= -\cos B_f \sin B_f \left(1 + 4{\eta_f}^2\right) \end{aligned}\right\}. \tag{6.42}$$

Inserting (6.40) and (6.42) into (6.41) gives:

$$B = B_f + {V_f}^2 \cos B_f (q - q_f) - \frac{1}{2}\cos B_f \sin B_f \left(1 + 4{\eta_f}^2\right)(q - q_f)^2. \tag{6.43}$$

According to the first equation in (6.38), we get:

$$\left.\begin{aligned} (q - q_f) &= -\frac{t_f \sec B_f}{2N_f^2} y^2 + \frac{t \sec B_f}{24N_f^4}\left(5 + 6t_f^2 + \eta_f^2\right) y^4 \\ (q - q_f)^2 &= \frac{t_f^2 \sec^2 B_f}{4N_f^4} y^4 \end{aligned}\right\}. \tag{6.44}$$

Inserting again into (6.43) and rearranging yields:

$$B = B_f - \frac{t_f}{2M_f N_f} y^2 + \frac{t_f}{24M_f N_f^3}\left(5 + 3t_f^2 + \eta_f^2 - 9\eta_f^2 t_f^2\right) y^4. \tag{6.45}$$

Combining (6.45) with the second equation in (6.38), the formula for the inverse solution of the Gauss projection can be obtained as follows:

$$\left.\begin{aligned}(B_f - B) &= \frac{t_f}{2M_f N_f} y^2 - \frac{t_f}{24 M_f N_f^3}\left(5 + 3t_f^2 + \eta_f^2 - 9\eta_f^2 t_f^2\right) y^4 \\ l &= \frac{1}{N_f \cos B_f} y - \frac{1}{6N_f^3 \cos B_f}\left(1 + 2t_f^2 + \eta_f^2\right) y^3 + \frac{1}{120 N_f^5 \cos B_f}\left(5 + 28t_f^2 + 24t_f^4\right) y^5\end{aligned}\right\}. \tag{6.46}$$

When $l < 3.5^\circ$, the accuracy of the computation according to (6.46) is $0.01''$. If the desired accuracy is $0.0001''$, the series in (6.46) can be further expanded:

$$\left.\begin{aligned}(B_f - B) &= \frac{t_f}{2M_f N_f} y^2 - \frac{t_f}{24 M_f N_f^3}\left(5 + 3t_f^2 + \eta_f^2 - 9\eta_f^2 t_f^2\right) y^4 \\ &\quad + \frac{t_f}{720 M_f N_f^5}\left(61 + 90t_f^2 + 45t_f^4\right) y^6 \\ l &= \frac{1}{N_f \cos B_f} y - \frac{1}{6N_f^3 \cos B_f}\left(1 + 2t_f^2 + \eta_f^2\right) y^3 \\ &\quad + \frac{1}{120 N_f^5 \cos B_f}\left(5 + 28t_f^2 + 24_f^4 + 6\eta_f^2 + 8\eta_f^2 t_f^2\right) y^5\end{aligned}\right\}. \tag{6.47}$$

L and B can finally be obtained according to:

$$\left.\begin{aligned}L &= L_0 + l \\ B &= B_f - (B_f - B)\end{aligned}\right\}. \tag{6.48}$$

Practical Formulae

Formulae for the Inverse Solution of the Gauss Projection (Accurate to $0.0001''$; The Results Are Measured in Degrees)

From (6.47), we have:

$$\left.\begin{aligned}B^\circ &= B_f^\circ - \frac{1}{2} V_f^2 t_f \left[\left(\frac{y}{N_f}\right)^2 - \frac{1}{12}\left(5 + 3t_f^2 + \eta_f^2 - 9\eta_f^2 t_f^2\right)\left(\frac{y}{N_f}\right)^4 \right. \\ &\quad \left. + \frac{1}{360}\left(61 + 90t_f^2 + 45t_f^4\right)\left(\frac{y}{N_f}\right)^6\right] \frac{180}{\pi} \\ l^\circ &= \frac{1}{\cos B_f}\left[\left(\frac{y}{N_f}\right) - \frac{1}{6}\left(1 + 2t_f^2 + \eta_f^2\right)\left(\frac{y}{N_f}\right)^3 \right. \\ &\quad \left. + \frac{1}{120}\left(5 + 28t_f^2 + 24t_f^4 + 6\eta_f^2 + 8\eta_f^2 t_f^2\right)\left(\frac{y}{N_f}\right)^5\right] \frac{180}{\pi}\end{aligned}\right\}, \tag{6.49}$$

where B_f is the footprint latitude; namely, the corresponding geodetic latitude with $x = X$ (the length of the meridian arc from the equator, where X corresponds to X_f). Footprint latitude B_f can be obtained by iteration. Equation (5.41) that corresponds to the Krassowski Ellipsoid is taken as an example to illustrate.

Given X, to compute B_f reversely, when iteration starts, the initial value is set:

$$B_f^{(1)} = X/111134.8611. \tag{6.50}$$

The ensuing iterative procedure follows:

$$B_f^{(i+1)} = \left(X - F\left(B_f^{(i)}\right)\right)/111134.8611, \tag{6.51}$$

$$F\left(B_f^{(i)}\right) = -\left(32005.7799 \sin B_f^{(i)} + 133.9238 \sin^3 B_f^{(i)} + 0.6973 \sin^5 B_f^{(i)} + 0.0039 \sin^7 B_f^{(i)}\right) \cos B_f^{(i)}. \tag{6.52}$$

Iterations are repeated until $|B_f^{(i+1)} - B_f^{(i)}| < 1 \times 10^{-8}$, to ensure that B_f is accurate to $0.0001''$. Normally, iterating five times will achieve the desired accuracy. It should be noted that in programming computation, B_f obtained using the iteration formula is measured in "degrees," whereas in iterative calculations, B_f in the trigonometric function is measured in "radians," so one needs to convert in iterative procedures.

Similarly, for GRS75 or GRS80 Ellipsoids, given X to compute B_f reservedly, one can iterate using (5.42) and (5.43), respectively.

Formulae for the Inverse Solution of the Gauss Projection (Accurate to $0.01''$; The Results Are Measured in Degrees)

It follows from (6.46) that:

$$\left.\begin{aligned} B^\circ &= B_f^\circ - \frac{1}{2} V_f^2 t_f \left[\left(\frac{y}{N_f}\right)^2 - \frac{1}{12}\left(5 + 3t_f^2 + \eta_f^2 - 9\eta_f^2 t_f^2\right)\left(\frac{y}{N_f}\right)^4\right]\frac{180}{\pi} \\ l^\circ &= \frac{1}{\cos B_f}\left[\left(\frac{y}{N_f}\right) - \frac{1}{6}\left(1 + 2t_f^2 + \eta_f^2\right)\left(\frac{y}{N_f}\right)^3 + \frac{1}{120}\left(5 + 28t_f^2 + 24t_f^4\right)\left(\frac{y}{N_f}\right)^5\right]\frac{180}{\pi} \end{aligned}\right\}, \tag{6.53}$$

where B_f can be obtained by iteration. The method is as stated above. The number of decimal places can be reduced taking into consideration the circumstances.

Table 6.2 Inverse solution of the Gauss projection: sample computation

		Computational results	
Given data (m)	Ellipsoidal parameters	6° zones	3° zones
$x = 3{,}354{,}874.257$	Krassowski Ellipsoid	$B = 30^\circ\ 18'\ 46.92''$	$B = 30^\circ\ 18'\ 46.92''$
$y = 386.564$		$L = 117^\circ\ 00'\ 14.47''$	$L = 60^\circ\ 00'\ 14.47''$
$y_{false} = 20, 500, 386.564$	GRS75 Ellipsoid	$B = 30^\circ\ 18'\ 48.80''$	$B = 30^\circ\ 18'\ 48.80''$
		$L = 117^\circ\ 00'\ 14.47''$	$L = 60^\circ\ 00'\ 14.47''$
	GRS80 Ellipsoid	$B = 30^\circ\ 18'\ 48.85''$	$B = 30^\circ\ 18'\ 48.85''$
		$L = 117^\circ\ 00'\ 14.47''$	$L = 60^\circ\ 00'\ 14.47''$
$x = 532\ 548.378$	Krassowski Ellipsoid	$B = 4^\circ\ 48'\ 57.62''$	$B = 4^\circ\ 48'\ 57.62''$
$y = -209.135$		$L = 116^\circ\ 59'\ 53.21''$	$L = 59^\circ\ 59'\ 53.21''$
$y_{false} = 20\ 499\ 790.865$	GRS75 Ellipsoid	$B = 4^\circ\ 48'\ 57.92''$	$B = 4^\circ\ 48'\ 57.92''$
		$L = 116^\circ\ 59'\ 53.21''$	$L = 59^\circ\ 59'\ 53.21''$
	GRS80 Ellipsoid	$B = 4^\circ\ 48'\ 57.93''$	$B = 4^\circ\ 48'\ 57.93''$
		$L = 116^\circ\ 59'\ 53.21''$	$L = 59^\circ\ 59'\ 53.21''$

To carry out computations, we use computer programming based on the methods and formulae above. The results of the sample computation are shown in Table 6.2, (assume that the central meridian in the 6° zone 20 is chosen as the central meridian in both 6° and 3° zones).

6.4.3 Transformation of Gauss Plane Coordinates Between Adjacent Zones

The Basics

To constrain distance distortions of the Gauss projection, the area of interest should be divided into zones along the meridians. The result of such a zone-dividing projection is that the unified coordinate system on the ellipsoid is divided into separate plane rectangular coordinate system in each zone. Hence, the points in adjacent zones belong to two coordinate systems; if transformed into the same coordinate system, the Gauss coordinates in one zone should be transformed into the coordinates in the neighboring zone, which is known as the Gauss coordinate transformation between adjacent zones.

In applications, transformation between adjacent zones is needed on the following occasions:

1. When geodetic control networks cross between different projection zones, coordinates of the adjacent zone should, partially or totally, be transformed into the same zone before adjustment computations are carried out.

2. While mapping in projection zone boundaries, the control points in adjacent zones are usually needed, which should be transformed into the same zone accordingly.
3. Zones 3° wide are needed for large-scale mapping (1:10,000 or larger-scale). However, the national control points usually only have the coordinates of the 6° zones. Hence, a transformation is also needed between the 6° zones and the 3° zones.

After deriving the formulae for direct and inverse solutions of the Gauss projection, transformation between adjacent zones will be easy. First, we compute the geodetic coordinates (L, B) of a point given the Gauss plane coordinates (x, y) of this point in zone I according to the formula for inverse solution of the Gauss projection. Then, we compute the Gauss plane coordinates $(x, y)_{\mathrm{II}}$ of this point in zone II based on the longitude of the central meridian $(L_0)_{\mathrm{II}}$ in this zone according to the formula for direct solution of Gauss projection. The process can be expressed as:

$$(x,y)_{\mathrm{I}} \xrightarrow[\text{inverse solution of the Gauss projection}]{(L_0)_{\mathrm{I}}} (L,B) \xrightarrow[\text{direct solution of the Gauss projection}]{(L_0)_{\mathrm{II}}} (x,y)_{\mathrm{II}}$$

Coordinate Transformation Between Zones with 3° Width and Zones with 6° Width

We know that the central meridian of 3° zones, in even-numbered zones, coincide with central meridian of the zones 6° wide, and in odd-numbered zones coincide with the zone-dividing meridian of the 6° zones. Hence, the coordinate transformation between the zones 3° and 6° wide has also to be considered in two different situations.

When the Central Meridian of the 3° Zone Coincides With That of the 6° Zone

In Fig. 6.10, the central meridian of the 3° zone 41 and that of the 6° zone 21 coincide, for both of which the longitude 123° is chosen as the central meridian. The difference between central meridians causes the coordinate system in each projection zone to be different. If the central meridians in the projection zones are coincident, then their coordinate systems will be the same. Given the coordinates of point P_1 in the 3° zone 41 to compute its coordinates in the 6° zone 21, no transformation is needed and vice versa.

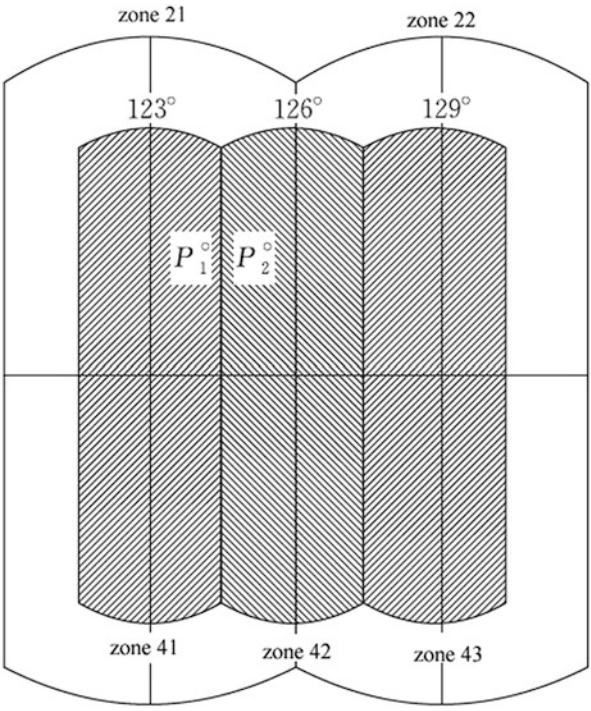


Fig. 6.10 Coordinate transformation between zones of 3° and 6° width

When the Central Meridian of the 3° Wide Zone Coincides With the Zone-Dividing Meridian of the 6° Wide Zone

In Fig. 6.10, the central meridian of the 3° zone 42 is coincident with the dividing meridian between zones 21 and 22 of the 6° zone. Their coordinate systems are different. Given the coordinates of point P_2 in the 3° zone 42 to compute its coordinates in the 6° zone 21, one can compute the coordinates of point P_2 in the 3° zone 41 according to the method of transformation between adjacent zones, and then its coordinates in 6° zone 21 can be obtained. The method used to transform coordinates of 6° zone into those of 3° zone is the same and will not be repeated here.

Example
Sample computations are provided in Table 6.3.

Table 6.3 Transformation between 3° zones and 6° zones: sample computations

	Central meridian of the zones 3° wide coincides with that of the zones 6° wide		Central meridian of the zones 3° wide coincides with the dividing meridian of the zones 6° wide	
Ellipsoidal parameters	Coordinates of 3° zone (m)	Coordinates of 6° zone (m)	Coordinates of 3° zone (m)	Coordinates of 6° zone (m)
Krassowski Ellipsoid	$x = 3{,}858{,}520.6946$	$x = 3{,}858{,}520.6946$	$x = 3{,}858{,}853.5671$	$x = 3{,}860{,}592.2479$
	$y_{\text{false}} = 41,\ 512,\ 354.9834$	$y_{\text{false}} = 21,\ 512,\ 354.9834$	$y_{\text{false}} = 42,\ 420,\ 902.8543$	$y_{\text{false}} = 21,\ 695,\ 272.9325$
GRS75 Ellipsoid		$x = 3{,}858{,}520.6946$		$x = 3{,}860{,}592.1771$
		$y_{\text{false}} = 21,\ 512,\ 354.9834$		$y_{\text{false}} = 21,\ 695,\ 266.4644$
GRS80 Ellipsoid		$x = 3{,}858{,}520.6946$		$x = 3{,}860{,}592.1751$
		$y_{\text{false}} = 21,\ 512,\ 354.9834$		$y_{\text{false}} = 21,\ 695,\ 266.2813$

6.5 Elements of the Geodetic Control Network Reduced to the Gauss Plane

6.5.1 Reduction of the Geodetic Control Network on the Ellipsoid to the Gauss Plane

According to the method for "reducing terrestrial observation elements to the ellipsoid" introduced in Chap. 5, we can reduce the elements of field observations (the horizontal direction, zenith distance, terrestrial distance, and astronomical azimuth, etc.) to the ellipsoid and then solve geodetic problems. Hence, the geodetic control network constituted by points on the Earth's surface can be reduced to the control network formed by geodetic points on the ellipsoid. However, the practical attempt to solve geodetic problems on the surface of the Earth ellipsoid has been proved remarkably complex. In addition, the control network on the ellipsoid cannot provide direct control for topographic mapping. In order to meet the needs of controlling topographic mapping and simplifying computations of the control network, we need to employ the method of Gauss projection to reduce the geodetic network constituted by geodesics on the ellipsoid to the geodetic network connected by straight lines on the plane. The computations entailed by the process of reduction are hereby presented specifically.

First, a few concepts are introduced below.

True North and Grid North

The so-called true north of a point refers to the direction of the northern end of the true meridian (geodetic meridian) which passes through the point, namely the direction towards the Earth's geographic North Pole. Grid north is the direction of the northern end of a straight line which passes through a point in the Gauss plane parallel to the vertical axis (northern direction of the north–south grid lines).

True Bearing and Grid Bearing (Azimuth)

True bearing is the angle between the true meridian and the geodesic, i.e., the geodetic azimuth. Grid bearing is the angle between the direction of the true north and that of a straight line on a plane. It is measured as positive clockwise from the grid north.

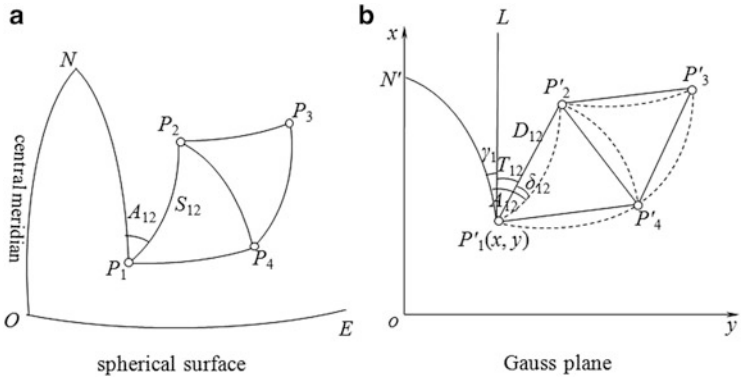


Fig. 6.11 Gauss projection of geodetic network from a spherical surface (**a**) to a plane (**b**). *Dashed lines* indicate projection of the sides of the triangles

Grid Convergence (Meridian Convergence)

The angular difference between grid north and true north is known as the grid convergence, denoted by γ. The grid convergence is measured from the projected meridian to the vertical grid lines, positive in the clockwise direction and negative in the anticlockwise direction.

In Fig. 6.11 we assume that a geodetic network $P_1P_2P_3P_4 \ldots$ is on the ellipsoid. The geodetic coordinates of the initial point P_1 is (L_1, B_1). Length of the initial side P_1P_2 (geodesic distance) is $S_{1.2}$. The initial geodetic azimuth is $A_{1.2}$. The observed values are those of the directions of each side of the triangulation network. After Gauss projection, the central meridian ON is projected as the axis of ordinates, i.e., the x-axis. The equator OE is projected as the axis of abscissas, i.e., the y-axis. The geodetic network $P_1P_2P_3P_4 \ldots$ is projected as $P_1'P_2'P_3'P_4' \ldots$ on the plane. According to the direct solution of the Gauss projection, all meridians, parallels, and geodesics are projected as curves except the central meridian and equator. Hence, the sides of the triangle formed by connecting geodesics on the ellipsoid are projected as the corresponding curves (expressed by dashed lines in Fig. 6.11). The meridian P_1N that passes through P_1 is $P_1'N'$ after projection, namely the direction of true north. A line $P_1'L$ drawn through P_1' parallel to the x-axis is the direction of grid north.

As a Gauss projection is conformal, the angles of a triangle on the ellipsoid will remain unchanged after being projected. Therefore, the angles constituted by the dashed lines on the Gauss plane are equal to the corresponding angles on the ellipsoid. The geodesics generally appear as curves rather than straight lines in the projection (i.e., the dashed lines in Fig. 6.11). To satisfy the needs of computations on the plane, one first needs to replace curves with chords that connect different points (represented by solid lines in Fig. 6.11). It is therefore necessary to

apply certain corrections to every direction, transforming the curves into the corresponding straight lines, and then to reduce the known elements in the geodetic network to the plane and convert the geodetic coordinates of the initial point to the plane coordinates. We compute the plane coordinates of other geodetic points, e.g., the plane coordinates of P_2' (Fig. 6.11b), with:

$$\begin{cases} x_2 = x_1 + D_{12} \cdot \cos T_{12} \\ y_2 = y_1 + D_{12} \cdot \sin T_{12} \end{cases},$$

One also needs to determine all the side lengths and grid azimuths of the plane triangle, such as D_{12} and T_{12}, etc.

From the above analysis, we can see that the reduction of the geodetic network on the ellipsoid to the Gauss plane consists of the following computations.

Direct Solution of the Gauss Projection

We reduce the geodetic coordinates (L_1, B_1) of the initial point P_1 to the Gauss plane rectangular coordinates (x_1, y_1) of the corresponding projection point P_1'. This is known as the direct solution of the Gauss projection.

Arc-to-Chord Correction

We convert the interior angles of the ellipsoidal triangle to that of the plane triangle formed by the corresponding straight lines. Actually, we convert the direction of the curve projected from a geodesic to the direction of its chord, and we obtain the angular difference between the projected curve and the chord, which is known as the arc-to-chord correction or curvature correction or direction correction, denoted by δ_{12}, δ_{13}, and so on.

Distance Correction

We reduce the geodesic distance $S_{1.2}$ of the initial side P_1P_2 on the ellipsoid to the length of chord $D_{1.2}$ of the corresponding projected side $P_1'P_2'$ on the plane, where the correction applied is called the correction of distance, denoted by ΔS.

Computation of Grid Convergence

We convert the geodetic azimuth $A_{1.2}$ of the initial side on the ellipsoid to the plane grid azimuth $T_{1.2}$ of the corresponding projected side $P_1'P_2'$, which is known as the

computation of grid azimuth, realized by reckoning the grid convergence γ and the arc-to-chord correction δ of this point.

After the above computations, the curved triangle on a plane constituted by curved lines will be transformed into the plane triangle formed by straight lines. As a result, the solution of a triangle and computation of plane coordinates can both be performed using the formulae for a plane triangle, which can greatly simplify geodetic computations.

Hence, to reduce the geodetic network on the ellipsoid to the plane, one needs to carry out computations such as coordinate transformation, arc-to-chord correction, correction of distance, computation of grid convergence and grid bearing, etc. Coordinate transformation is one of the direct and inverse solutions of the Gauss projection, which were introduced in previous sections. Next, our discussions will focus on other computations.

6.5.2 Arc-to-Chord Correction

The correction applied to compensate for the distortion of a straight line when a geodesic between two points on the ellipsoid is reduced to the chord between corresponding projection points on the plane is called the arc-to-chord correction, denoted by δ_{ij}. As the Gauss projection is conformal, the direction of the geodesic remains unchanged after being projected. Hence, the arc-to-chord correction can also be interpreted as a process of converting the curve projected from a geodesic to the chord between the corresponding two points, i.e., the angle between the projected curve and the chord. Such an angle exists owing to distortions of the geodesic curves after being projected. Its magnitude depends on the curvature of the curve, which is therefore also referred to as the curvature correction. It can be seen that the need for arc-to-chord correction is caused by the fact that, on the plane, the curves are projected as straight lines, rather than by the distortion of projection.

The accuracy and form of the formulae for the arc-to-chord correction in the Gauss projection vary with the order of computation. Typically, precise formulae are applied to the first-order correction, relatively precise formulae are applied to the second-order net, and for the third- and fourth-order calculations we use the approximate formulae.

Approximation Formula for Arc-to-Chord Correction

As shown in Fig. 6.12a, the ellipsoid is approximated as a sphere, and the geodesic P_1P_2 will be the great circle (orthodrome) on the spherical surface. We draw two great circles AP_1 and BP_2 passing through points P_1 and P_2 perpendicular to the central meridian, and both of them intersect the equator at point E. ABP_2P_1 constitutes a spherical quadrangle. In Fig. 6.12b the geodesic P_1P_2 is projected as the curve $P_1'P_2'$. As Gauss projection is conformal, the great circles AP_1 and BP_2

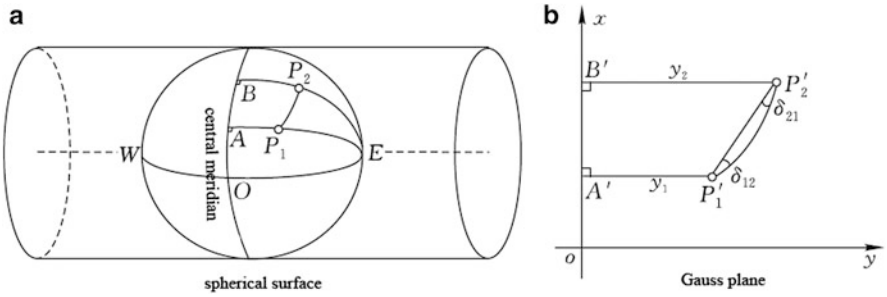


Fig. 6.12 Approximate derivations of the arc-to-chord correction for the Gauss projection from a spherical surface (**a**) to a plane (**b**)

appear as straight lines $A'P'_1$ and $B'P'_2$ in the projection. They are both perpendicular to the x-axis (for point E is projected to an infinite distance). $A'B'P'_2P'_1$ constitutes a plane-curved quadrangle, where the side $P'_1P'_2$ is a curved line.

Let the spherical excess of the spherical quadrangle be ε and the angles between the projected geodesic $P'_1P'_2$ and its chord $\overline{P'_1P'_2}$ be $\delta_{1.2}$ and $\delta_{2.1}$. For the projection being conformal, the angle relationship is:

$$360^\circ + \varepsilon = 360^\circ + \delta_{1.2} + \delta_{2.1} \tag{6.54}$$

Let $\delta_{1.2} = \delta_{2.1} = \delta$; then one obtains:

$$\delta = \frac{\varepsilon}{2},$$

where $\varepsilon = \frac{P}{R^2}$ and P is the area of the spherical quadrangle. Since the numerical value of ε is very small, P can be replaced by the area of the plane quadrangle. We assume that the plane coordinates of P'_1 and P'_2 are (x_1, y_1) and (x_2, y_2); thus:

$$P = \frac{1}{2}(y_1 + y_2)(x_2 - x_1) = y_m(x_2 - x_1),$$

and

$$\delta = \frac{y_m}{2R_m^2}(x_2 - x_1), \tag{6.55}$$

where R_m is the mean radius of curvature at the mid-latitude B_m of the two end points P_1 and P_2.

The above derivations only result in the absolute values of the arc-to-chord correction. However, in practical cases, as the position and direction of geodesics vary, the value of δ can be either positive or negative. To enable the corrections to appear in the form of the algebraic sum, make the δ obtained be the correction

Table 6.4 Datasheet for the arc-to-chord correction (δ''_{12}, in seconds) for different side lengths (x_2-x_1) and distances (y_m) of the side from the central meridian

	x_2-x_1 (km)										
y_m (km)	0	4	8	12	16	20	24	28	32	36	40
100	0.0	1.0	2.0	3.0	4.0	5.1	6.1	7.1	8.1	9.1	10.1
200	0.0	2.0	4.1	6.1	8.1	10.1	12.2	14.2	16.2	18.3	20.3
300	0.0	3.0	6.1	9.1	12.2	15.2	18.2	21.3	24.3	27.4	30.4

applied to the observed direction. Note that the sign of δ should be taken into account. For instance, in Fig. 6.12b, as the values of observed directions increase in the clockwise direction, when converting the geodesic P_1P_2 to its equivalent chord $\overline{P'_1P'_2}$, the sign of the arc-to-chord correction δ_{12} is negative. Similarly, when converting P_2P_1 to $\overline{P'_2P'_1}$, the sign of the arc-to-chord correction δ_{12} becomes positive. Represented by seconds, the approximation formula for the arc-to-chord correction is:

$$\left.\begin{aligned}\delta''_{1.2} &= -\frac{\rho'' y_m}{2R_m^2}(x_2 - x_1)\\ \delta''_{2.1} &= \frac{\rho'' y_m}{2R_m^2}(x_2 - x_1)\end{aligned}\right\}. \tag{6.56}$$

This formula has an accuracy of better than $0.1''$ and is typically applied to computations of the third-order triangulation or lower.

It can be seen from (6.56) that the value of an arc-to-chord correction is likely to become larger the further away the side is from the central meridian. Some numeric values of the arc-to-chord correction calculated according to (6.56) are listed in Table 6.4.

In Table 6.4, (x_2-x_1) roughly corresponds to the side length of the geodetic network and y_m is approximately the distance of the side from the central meridian. It becomes obvious that the arc-to-chord corrections are not negligible for triangulations of various orders.

Precise Formula for Arc-to-Chord Correction

In much of the literature, the approximate formulae are used as differential equations to derive the relatively precise formulae. The new coordinate systems and formulae for radii of curvature are introduced to establish the second-order differential equation, and then the solution is found by solving the differential equation. The derivation is rather long-winded and therefore will not be discussed in detail here.

The relatively precise formulae for the arc-to-chord correction are:

$$\left.\begin{aligned}\delta''_{1.2} &= -\frac{\rho''}{6{R_m}^2}(x_2 - x_1)(2y_1 + y_2)\\ \delta''_{2.1} &= \frac{\rho''}{6{R_m}^2}(x_2 - x_1)(2y_2 + y_1)\end{aligned}\right\}. \tag{6.57}$$

The average side-length of China's second-order triangulation network is 13 km. When y_m is less than 250 km, the above equations are accurate to 0.01 " and are typically used in computations of second-order triangulation. When y_m is greater than 250 km, we should apply the precise formulae for arc-to-chord correction in (6.58):

$$\left.\begin{aligned}\delta''_{1.2} &= -\frac{\rho''}{6{R_m}^2}(x_2 - x_1)\left(2y_1 + y_2 - \frac{{y_m}^3}{{R_m}^2}\right) - \frac{\rho''{\eta_m}^2 t_m}{R_m^3}(y_2 - y_1){y_m}^2\\ \delta''_{2.1} &= \frac{\rho''}{6{R_m}^2}(x_2 - x_1)\left(2y_2 + y_1 - \frac{{y_m}^3}{{R_m}^2}\right) + \frac{\rho''{\eta_m}^2 t_m}{R_m^3}(y_2 - y_1){y_m}^2\end{aligned}\right\}. \tag{6.58}$$

The above formulae are accurate to $0.001''$ and are applicable to computations of the first-order triangulation.

Accuracy of Coordinates Required in Computations of Arc-to-Chord Correction

To calculate the arc-to-chord correction, we should first obtain the plane coordinates of a point. Paradoxically, knowing precisely the plane coordinates of a point also requires that the arc-to-chord correction be computed first. The way to resolve this contradiction is to apply the iterative computing method. As computations of different orders require different degrees of accuracy, the number of iterations is also different. Here we will analyze the accuracy of coordinates required.

Taking the total differential of (6.56) gives:

$$\Delta\delta'' = \frac{\rho''}{2{R_m}^2}[y_m \cdot \Delta(x_2 - x_1) + (x_2 - x_1) \cdot \Delta y].$$

We set $\Delta(x_2 - x_1) = \Delta y = \Delta P$, to obtain:

$$\Delta\delta'' = \frac{\rho''}{2{R_m}^2} \cdot \Delta P \cdot [y_m + (x_2 - x_1).$$

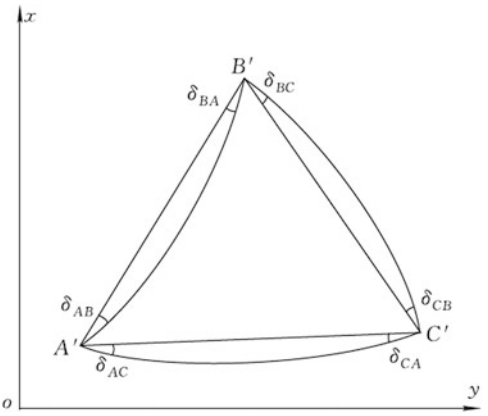


Fig. 6.13 Arc-to-chord correction checking

Rearranging gives:

$$\Delta P = \frac{2{R_m}^2}{\rho''} \frac{\Delta\delta''}{y_m + (x_2 - x_1)}.$$

In the third-order triangulation we require that $\Delta\delta'' = 0.\ 1''$. With $y = 350$ km and $x_2 - x_1 = 10$ km we get $\Delta P \approx 0.1$ km. It follows that the approximate coordinates should be accurate to 0.1 km to meet the desired accuracy in computations of the third-order arc-to-chord correction. Likewise, for the first- and second-order arc-to-chord correction, the plane coordinates should be accurate to 10 m and 1 m, respectively. For many third-order triangulations, the desired accuracy of approximate coordinates is not high and thus iterative computation is not necessary.

Formula for Checking the Computation of Arc-to-Chord Correction

The sum of the measures of the interior angles of an ellipsoidal triangle is $180^\circ + \varepsilon$, and that of the triangle formed by curves remains unchanged after the ellipsoidal triangle is conformally projected onto a plane. As shown in Fig. 6.13, the curved triangle $A'B'C'$ is the projection of the ellipsoidal triangle ABC onto the Gauss plane. Let the angular correction of each angle be δ_A, δ_B, δ_C, which are equal to the differences in the arc-to-chord correction between the two neighboring sides, namely:

$$\left.\begin{aligned} \delta_A &= \delta_{AC} - \delta_{AB} \\ \delta_B &= \delta_{BA} - \delta_{BC} \\ \delta_C &= \delta_{CB} - \delta_{CA} \end{aligned}\right\}. \tag{6.59}$$

The interior angles of an ellipsoidal triangle add up to $180^\circ + \varepsilon$. After angular corrections are applied, the ellipsoidal triangle becomes a triangle formed by straight line sides with the sum of interior angles 180°, namely:

$$180^\circ + \varepsilon + (\delta_A + \delta_B + \delta_C) = 180^\circ.$$

Then the spherical excess is:

$$\varepsilon = -(\delta_A + \delta_B + \delta_C) \tag{6.60}$$

The above expression shows that the spherical excess of each triangle is equal to the sum of the angular corrections of the interior angles of this plane triangle, of opposite signs. Equation (6.60) can be used to check the correctness of the arc-to-chord correction and computations of the spherical excess.

Practical Formulae

The formulae accurate to $0.001''$ are:

$$\left.\begin{aligned} \delta''_{1.2} &= -\frac{\rho''}{6{R_m}^2}\left[(x_2 - x_1)\left(2y_1 + y_2 - \frac{{y_m}^3}{{R_m}^2}\right) + \frac{6{\eta_m}^2 t_m}{R_m}(y_2 - y_1){y_m}^2\right] \\ \delta''_{2.1} &= -\delta''_{1.2} + \frac{\rho''}{6{R_m}^2}(x_2 - x_1)(y_2 - y_1) \end{aligned}\right\}. \tag{6.61}$$

For instance, given that $x_1 = 3602547.8$ m, $y_1 = 298960.0$ m, $x_2 = 3584223.0$ m, $y_2 = 323655.4$ m and $B_m = 32^\circ 25.5'$, the results give $\delta''_{1.2} = +14.294''$, and $\delta''_{2.1} = -14.678''$.

We check:

$$\sum \delta = \delta_1 + \delta_2 + \delta_3$$

$$= (16.430'' - 14.294'') + (-14.678'' - 2.577'') + (-2.523'' + 16.519'') = -1.123'',$$

and $\varepsilon = +1.123''$.

The two results have the same absolute value, indicating that the computation is correct, as shown in Fig. 6.14. Owing to the round-off errors, there would presumably be a difference of 0.001 "–0.002 " between them. After being proved errorless, the value of difference can be assigned to the large angle.

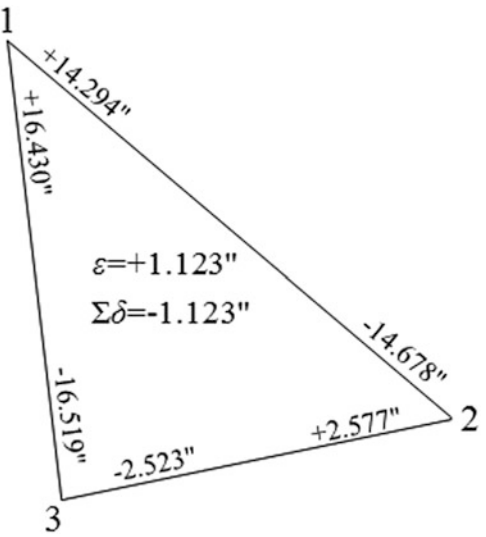


Fig. 6.14 Example of arc-to-chord correction checking

6.5.3 Correction of Distance

The Gauss projection is conformal with no angular distortion. However, the projection distorts distances everywhere except along the central meridian. The distance correction of the Gauss projection is dependent on the distortion of distance. Previously, we provided the definitions of the scale and distance distortion. Below are the derivations of the specific mathematical expressions for the scale so that the laws and effects of distance distortion and the ways to limit such distortions can be studied. The formula for the distance distortion will be derived after that.

Formula for Scale Factor

It has been previously stated that the ratio of the arc element at a point on the projection plane ds to the corresponding arc element on the ellipsoid dS is known as the scale factor at this point, namely $m = \dfrac{\mathrm{d}s}{\mathrm{d}S}$. For conformal projections, the scale factor at any point is independent of direction. Hence, any arbitrary directions can be selected when the formula for the scale factor is derived. In (6.16) we provided the formulae for scale factors along two special directions, where the first expression is that along the meridian (l = constant) and the second along the parallel (q = constant). In conjunction with the formula for the direct solution of the Gauss projection, it is reasonably convenient to take the partial derivative with respect to l. Therefore, it is simpler to use the second expression in (6.16) for deriving the formula for the scale factor, namely:

$$m^2 = \frac{G}{r^2} = \frac{\left(\frac{\partial x}{\partial l}\right)^2 + \left(\frac{\partial y}{\partial l}\right)^2}{N^2 \cos^2 B}. \tag{6.62}$$

Using the Geodetic Coordinates (B, l) to Derive the Formula for m

We take the partial derivative with respect to l by applying the formula for the direct solution of the Gauss projection (6.28) to get:

$$\left.\begin{aligned} \frac{\partial x}{\partial l} &= N \sin B \cos B l + \frac{N}{6} \sin B \cos^3 B \left(5 - t^2 + 9\eta^2 + 4\eta^4\right) l^3 \\ &\quad + \frac{N}{120} \sin B \cos^5 B \left(61 - 58t^2 + t^4\right) l^5 \\ \frac{\partial y}{\partial l} &= N \cos B + \frac{N}{2} \cos^3 B \left(1 - t^2 + \eta^2\right) l^2 \\ &\quad + \frac{N}{24} \cos^5 B \left(5 - 18t^2 + t^4 + 14\eta^2 - 58\eta^2 t^2\right) l^4 \end{aligned}\right\}. \tag{6.63}$$

In (6.63), we divide the obtained $\frac{\partial x}{\partial l}$ and $\frac{\partial y}{\partial l}$ (with the terms containing l^5 and $\eta^2 l^3$ left out) by $N \cos B$; then the square is:

$$\frac{\left(\frac{\partial x}{\partial l}\right)^2}{N^2 \cos^2 B} = l^2 \sin^2 B + \frac{l^4}{3} \sin^2 B \cos^2 B \left(5 - t^2\right)$$

$$\begin{aligned} \frac{\left(\frac{\partial y}{\partial l}\right)^2}{N^2 \cos^2 B} &= \left[1 + \frac{l^2}{2} \cos^2 B \left(1 - t^2 + \eta^2\right) + \frac{l^4}{24} \cos^4 B \left(5 - 18t^2 + t^4\right)\right]^2 \\ &= 1 + l^2 \cos^2 B (1 - t^2 + \eta^2) + \frac{l^4}{4} \cos^4 B \left(1 - t^2\right)^2 + \frac{l^4}{12} \cos^4 B \left(5 - 18t^2 + t^4\right) \\ &= 1 + l^2 \cos^2 B (1 - t^2 + \eta^2) + \frac{l^4}{3} \cos^4 B \left(2 - 6t^2 + t^4\right). \end{aligned}$$

Substituting into (6.62) yields:

$$\begin{aligned} m^2 &= 1 + l^2\cos^2 B(1 - t^2 + \eta^2) + l^2\sin^2 B + \frac{l^4}{3}\cos^4 B\left(2 - 6t^2 + t^4\right) \\ &\quad + \frac{l^4}{3}\sin^2 B\cos^2 B\left(5 - t^2\right) \\ &= 1 + l^2\cos^2 B(1 + \eta^2) + \frac{l^4}{3}\cos^4 B\left(2 - t^2\right) \\ m &= \left[1 + l^2\cos^2 B(1 + \eta^2) + \tfrac{l^4}{3}\cos^4 B(2 - t^2)\right]^{1/2}. \end{aligned}$$

The above is the approximate formula used to compute the scale factor m from given geodetic coordinates (B, l). If one more terms in the equation are allowed for in-formula derivations, we will then be able to achieve a formula for the scale m with a higher level of accuracy. Expanding the above equation according to the binomial theorem $(1 + x)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \ldots$ gives:

$$\begin{aligned} m &= 1 + \frac{1}{2}\left[l^2\cos^2 B\left(1 + \eta^2\right) + \frac{l^4}{3}\cos^4 B\left(2 - t^2\right)\right] - \frac{l^2}{8}\left[\cos^2 B\left(1 + \eta^2\right)\right]^2 \\ &= 1 + \frac{l^2}{2}\cos^2 B\left(1 + \eta^2\right) + \frac{l^4}{6}\cos^4 B\left(2 - t^2\right) - \frac{l^4}{8}\cos^4 B \\ &= 1 + \frac{l^2}{2}\cos^2 B\left(1 + \eta^2\right) + \frac{l^4}{24}\cos^4 B\left(5 - 4t^2\right). \end{aligned}$$

If l is measured in seconds, then we have:

$$m = 1 + \frac{l''^2}{2\rho''^2}\cos^2 B\left(1 + \eta^2\right) + \frac{l''^4}{24\rho''^4}\cos^4 B\left(5 - 4t^2\right), \tag{6.64}$$

which is the formula for the scale factor expressed by geodetic coordinates.

To Derive m Given Gauss Plane Coordinates (x, y)

According to the second equation in (6.28), and leaving out the m^5 term, we obtain:

$$y = N\cos Bl + \frac{N}{6}\cos^3 B\left(1 - t^2 + \eta^2\right)l^3 = Nl\cos B\left[1 + \frac{l^2}{6}\cos^2 B\left(1 - t^2 + \eta^2\right)\right].$$

Rearranging gives:

$$l\cos B = \frac{y}{N}\left[1 + \frac{l^2}{6}\cos^2 B\left(1 - t^2 + \eta^2\right)\right]^{-1} = \frac{y}{N}\left[1 - \frac{l^2}{6}\cos^2 B\left(1 - t^2 + \eta^2\right)\right].$$

We replace $l\cos B$ on the right-hand side of the above equation with $\frac{y}{N}$, which gives:

$$\frac{l''}{\rho''}\cos B = \frac{y}{N} - \frac{y^3}{6N^3}\left(1 - t^2 + \eta^2\right),$$
$$\frac{l''^2}{\rho''^2}\cos^2 B = \frac{y^2}{N^2} - \frac{y^4}{3N^4}\left(1 - t^2 + \eta^2\right),$$

and

$$\frac{l''^4}{\rho''^4}\cos^4 B = \frac{y^4}{N^4}.$$

Substituting into (6.64), one obtains:

$$m = 1 + \frac{1}{2}\left[\frac{y^2}{N^2} - \frac{y^4}{3N^4}\left(1 - t^2 + \eta^2\right)\right]\left(1 + \eta^2\right) + \frac{y^4}{24N^4}\left(5 - 4t^2\right).$$

It follows that:

$$m = 1 + \frac{y^2}{2N^2}\left(1 + \eta^2\right) + \frac{y^4}{24N^4}. \tag{6.65}$$

With $\frac{1}{R^2} = \frac{y^2}{N^2} = \frac{1}{N^2}(1 + \eta^2)$, substituting into the above equation and replacing N^4 with R^4, we get the formula for the scale factor expressed by the Gauss plane coordinates:

$$m = 1 + \frac{y^2}{2R^2} + \frac{y^4}{24R^4}. \tag{6.66}$$

Table 6.5 provides some approximate values for the scale factor.

Distortion Properties

Subtracting 1 from the point scale factor, i.e., $(m - 1)$ results in the distortion of distance at this point. Equation (6.66) shows that the point scale factor m depends only on position, not on direction (which agrees with the conditions for the

Table 6.5 Numerical values for the scale factor (m) for different distances (y) of the point from the central meridian at different latitudes (B) (GRS80 Ellipsoid)

y (km)	$B = 0°$	$B = 10°$	$B = 20°$	$B = 30°$	$B = 40°$
50	1.000030936	1.000030923	1.000030887	1.000030832	1.000030764
100	1.000123745	1.000123695	1.000123550	1.000123328	1.000123057
200	1.000495011	1.000494810	1.000494231	1.000493344	1.000492258
300	1.001113890	1.001113437	1.001112134	1.001110139	1.001107694
350	1.001516230	1.001515613	1.001513839	–	–
y (km)	$B = 50°$	$B = 60°$	$B = 70°$	$B = 80°$	$B = 90°$
50	1.000030692	1.000030624	1.000030569	1.000030533	1.000030520
100	1.000122768	1.000122497	1.000122277	1.000122133	–
200	1.000491103	1.000490019	–	–	–

conformal projection). With $y = 0$, $m = 1$, i.e., the central meridian remains unchanged in length after the projection (which is consistent with the conditions for the Gauss projection). In the case of $y \neq 0$, with its value being either positive or negative, m is invariably greater than 1, which means that the differential line segments away from the central meridian are all stretched after the projection. The distortion of distance $(m - 1)$ increases proportionally with y^2. For any arbitrary meridian, the distance distortion increases with distance from the standard parallels. Again, for any arbitrary meridian except for the central meridian, the distortion of distance is maximal at the equator, i.e., the lower the latitude, the more significant the distortion of distance.

The distortion of distance is harmful, but it actually exists. We cannot violate this law to eliminate the distortion. Therefore, in practical cases we can only impose appropriate limitations on distance distortion, making its effect on mapping and application of maps insignificant. The way to limit such a distortion is the zone-dividing projection.

As mentioned earlier, the projection zones in China primarily consist of 6° zones and 3° zones. The national basic maps at scales ranging from 1:25,000 to 1:100,000 use the zones 6° of longitude in width, and the topographic maps at scales of 1:10,000 or larger use the zones that are 3° of longitude wide. It is often the case that the distortion of distance is fairly significant in lower-latitude regions, so at north latitude 20° and further south in the south of China this effect in mapping and map application should be considered. For instance, it can be seen from (6.66) that $m - 1 = \frac{y^2}{2R^2}$; so, in areas at the north latitude 20°, close to the dividing meridian of 6° zones, the distortion of distance $m - 1$ can amount to $\frac{1}{820}$, which is quite large and cannot be neglected for the 1:25,000 or 1:50,000 scale maps. For example, in such a case, the side length of 10 km is subject to a distance distortion of 12.2 m. On the 1:25,000 scale map, such distortion is approximately 0.5 mm. Generally, the point position error on the map (i.e., mapping error) is allowed to be less than 0.2 mm. Such effects in mapping and map application must be taken into consideration.

For the 3° wide zone, in areas at latitude 20° or further south, the distance distortion of the 3° zone boundary reaches 1/3,300. Such a distortion still cannot be

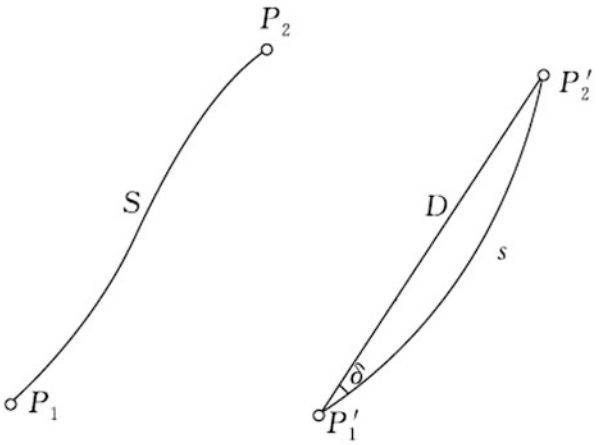


Fig. 6.15 Distance correction in converting geodesic distance S (*left*) to s and D on the Gauss plane (*right*)

overlooked for mapping and the use of maps at a scale of 1:5,000 or larger, and the corresponding corrections, is therefore necessary. One can also use either the 1.5° zone or an urban independent Gauss rectangular coordinate system (i.e., select the meridian passing through the center of a city as the central meridian) to allow the distortion of distance to satisfy the mapping needs.

Formula for Distance Correction

Derivations of Formula

In Fig. 6.15 we assume that S is the geodesic distance between two points P_1 and P_2 on the ellipsoid, s is the length of the projected curve between the corresponding points P'_1 and P'_2 projected on the Gauss plane, and D is the chord length between the two points P'_1 and P'_2 on the projected curve.

The correction added while converting the geodesic distance S to the plane chord D is known as the distance correction, denoted by ΔS.

Generally, the scale factor in a Gauss projection is invariably greater than 1, hence it follows that:

$$S < s > D$$

We are aiming to obtain the relationship between S and D. In the process of deduction, we will first approach the relationship between S and s, and then that between s and D. Finally, we will derive the formula for computing the distance correction ΔS.

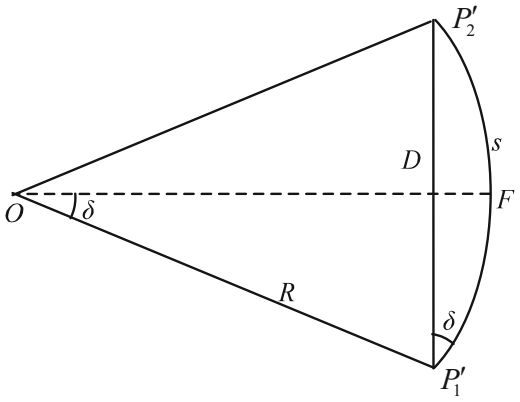


Fig. 6.16 Relationship between s and D

From the definition of the scale factor, we have:

$$\mathrm{d}s = m\mathrm{d}S.$$

Integrating produces:

$$s = \int_{P_1}^{P_2} m\mathrm{d}S = \int_0^S m\mathrm{d}S. \tag{6.67}$$

In practice, sometimes the integrals of some functions are hard to obtain, and in such cases we can approximate a definite integral according to the desired computational accuracy. In (6.67), m changes with the position of the point. When the projection area is not large, this change is slow. For instance, when $y = 300$ km, and the difference in latitude between points P_1 and P_2 reaches 1°, the difference in scale between the two points is less than 4×10^{-7}. Therefore, by approximating integrals, we can obtain satisfactory accuracy. Now, we shall find the solution of (6.67) according to Simpson's rule for numerical integration.

Simpson's rule for approximate integration uses parabolic arcs to replace the area bounded by curve $y = f(x)$ over an interval $[x_1, x_2]$ so as to find the definite integral. This parabola passes through three points, i.e., $y_1 = f(x_1)$, $y_2 = f(x_2)$, $y_0 = f(x_0)$, with $x_0 = \dfrac{x_1 + x_2}{2}$, dividing the interval $[x_1, x_2]$ into two subintervals. It is easy to obtain the definite integral of the parabola on the interval $[x_1, x_2]$, which is $\dfrac{(x_2 - x_1)}{6}(y_1 + 4y_0 + y_2)$. Taking the definite integral of this parabola as the approximation to the integral of $f(x)$, we have:

$$\int_{x_1}^{x_2} f(x)\mathrm{d}x = \frac{(x_2 - x_1)}{6}(y_1 + 4y_0 + y_2).$$

This is Simpson's rule for the approximate integration.

According to Simpson's rule for integration, we divide the interval of integration in (6.67) into two subintervals of equal width $\frac{S}{2}$; then:

$$s = \frac{S}{6}(m_1 + 4m_m + m_2), \tag{6.68}$$

where m_1 and m_2 are the scale factors at points P_1 and P_2, respectively. m_m is the scale factor at the midpoint of the geodesic.

Now we will derive the relationship between s and D. As shown in Fig. 6.16, $P_1'P_2'$ is the projected curve of geodesic P_1P_2, approximating a circular arc, where O is the center of the circular arc, F is the midpoint, and δ is the curvature correction. Hence $\angle\, P_1'OF$ is also δ. The relationship is given by:

$$\sin\delta = \frac{D/2}{R} = \frac{D}{2R}, \quad R = \frac{s}{2\delta}.$$

It follows from the above two equations that:

$$D = \frac{s\sin\delta}{\delta},$$

and $\sin\delta = \delta - \frac{\delta^3}{3!} + \frac{\delta^5}{5!} - \frac{\delta^7}{7!} + \cdots,$

with $\delta = 30\ '' = 0.00015$ rad, $s = 40$ km, and $\frac{\delta^2}{3!}s = 0.15$ mm, the required side lengths of the first-order triangulation are defined to maintain an accuracy to millimeter level, so this term and higher-order terms can all be neglected. As a result, in computations, one can consider:

$$D = s. \tag{6.69}$$

We replace s in (6.68) with D to get:

$$D = \frac{S}{6}(m_1 + 4m_m + m_2). \tag{6.70}$$

Again, according to (6.66), we have:

$$m_1 = 1 + \frac{{y_1}^2}{2{R_1}^2} + \frac{{y_1}^4}{24{R_1}^4}$$

$$m_m = 1 + \frac{{y_m}^2}{2{R_m}^2} + \frac{{y_m}^4}{24{R_m}^4}$$

$$m_2 = 1 + \frac{{y_2}^2}{2{R_2}^2} + \frac{{y_2}^4}{24{R_2}^4}.$$

Inserting into (6.70) and replacing $\frac{1}{R_1^2}$ and $\frac{1}{R_2^2}$ with $\frac{1}{R_m^2}$ produces:

$$D = \frac{S}{6}\left(6 + \frac{{y_1}^2}{2{R_m}^2} + 4\frac{{y_m}^2}{2{R_m}^2} + \frac{{y_2}^2}{2{R_m}^2} + \frac{{y_1}^4}{24{R_m}^4} + 4\frac{{y_m}^4}{24{R_m}^4} + \frac{{y_2}^4}{24{R_m}^4}\right).$$

We set

$$y_m = \frac{y_1 + y_2}{2}, \qquad \frac{\Delta y}{2} = \frac{y_2 - y_1}{2}$$

to obtain

$$y_1 = y_m - \frac{\Delta y}{2}, \quad y_2 = y_m + \frac{\Delta y}{2}$$

and ${y_1}^2 + {y_2}^2 = 2{y_m}^2 + \frac{\Delta y^2}{2}$.

Since the term containing y^4 is minute, and hence replaced by ${y_m}^4 = \frac{{y_1}^4 + {y_2}^4}{2}$, one obtains:

$$D = S\left(1 + \frac{{y_m}^2}{2{R_m}^2} + \frac{\Delta y^2}{24{R_m}^2} + \frac{{y_m}^4}{24{R_m}^4}\right), \tag{6.71}$$

and the resulting equation:

$$\Delta S = D - S = S\left(\frac{{y_m}^2}{2{R_m}^2} + \frac{\Delta y^2}{24{R_m}^2} + \frac{{y_m}^4}{24{R_m}^4}\right). \tag{6.72}$$

The above is the formula for the distance correction of the Gauss projection. When $S < 70$ km and $y_m < 350$ km, the equation has an accuracy of 0.001 m, or better. Therefore, this formula is applicable to computations of the first-order triangulation.

For the second-order triangulation (accurate to 0.01 m), we can leave out the last term in the above equation, namely:

$$\Delta S = D - S = S\left(\frac{{y_m}^2}{2{R_m}^2} + \frac{\Delta y^2}{24{R_m}^2}\right). \tag{6.73}$$

For the third-order triangulation (accurate to 0.1 m), one has only to consider the first term:

$$\Delta S = D - S = \frac{{y_m}^2}{2{R_m}^2} S. \tag{6.74}$$

Thus, it can be seen that with $y = 0$ and $\Delta S = 0$, the distance correction on the central meridian is zero. In the case of $y \neq 0$ and $\Delta S > 0$, we see that the distance correction is always a positive value, increasing with the distance from the central meridian. When $y_m = 300$ km, $S = 5$ km, and $R_m = 6,400$ km, we can get $\Delta S = 6$ m after the distance correction. Obviously, the distance correction is non-negligible in computations of all orders.

Also, if converting the planar chord-length D to the geodesic distance S, we have:

$$S = D - \Delta S \tag{6.75}$$

Accuracy of Coordinates Required in the Computation of Distance Correction

To calculate the distance correction one needs to know the plane coordinates of a point. As the value of the distance correction is not large, it does not require too high a level of coordinate accuracy. To know the approximations of coordinates is sufficient; we will analyze the desired degree of accuracy for coordinates.

From (6.74), we get:

$$\Delta(D - S) = \frac{2y_m}{2{R_m}^2} S \Delta y$$

and

$$\Delta y = \frac{{R_m}^2}{y_m S} \Delta(D - S).$$

With $y_m = 350$ km, $S = 50$ km, and $R_m = 6,400$ km, and for the first-order triangulation, let $\Delta(D - S) = 0.001$ m; we obtain $\Delta y = 2.34$ m. In the meantime, for the second- and third-order triangulations, Δy is 23.4 m and 234 m, respectively. Hence, the coordinates with an accuracy of 1 m and 10 m would satisfy the first-, second-, and third-order triangulations. In a large number of computations, to avoid accumulative errors of coordinates, the coordinates are always accurate to 0.1 m, 1 m, and 10 m. Based on the analyses above, such an accuracy of coordinates can also satisfy the computation requirements for the arc-to-chord correction of the same order.

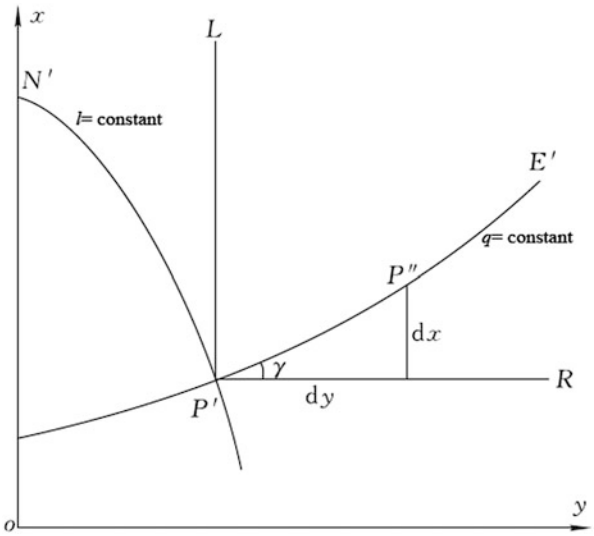


Fig. 6.17 Grid convergence

Practical Formula

Equation (6.71) is the practical formula that yields an accuracy of 0.001 m:

$$D = S\left[1 + \frac{1}{2}\left(\frac{y_m}{R_m}\right)^2 + \frac{1}{24}\left(\frac{\Delta y}{R_m}\right)^2 + \frac{1}{24}\left(\frac{y_m}{R_m}\right)^4\right], \tag{6.76}$$

where $R_m = \dfrac{c}{1 + e'^2 \cos^2 B_m}$; see (5.35). If B_m is unknown, we can put $X = \frac{1}{2}(x_1 + x_2)$ and calculate B_m according to the formula for computing latitude from the meridian arc length.

For instance, given $y_1 = 269759.6$ m, $y_2 = 297219.7$ m, $B_m = 31^\circ\ 27'$, and $S = 34862.820$ m, one obtains $D = 34{,}897.394$ m.

Alternatively, given $x_1 = 3496205.1$ m, $y_1 = 269759.6$ m, $x_2 = 3474669.9$ m, $y_2 = 297219.7$ m, and $S = 34862.820$ m, one gets $D = 34{,}897.394$ m.

6.5.4 Grid Convergence

Grid convergence is needed in determining the grid azimuth. We will hereby derive the formula for computing the grid convergence.

On the Gauss projection plane, as shown in Fig. 6.17, the angle between the projected meridian $P'N'$ passing through point P' (the curve with $l =$ constant) and the vertical grid line $P'L$ is referred to as the Gauss plane grid convergence. Because the Gauss projection is a conformal projection, the projected meridian and parallel

(i.e., $P'N'$ and $P'E'$) are still perpendicular to each other. Hence, the angular difference between $P'E'$ and the horizontal grid line $P'R$ is also the grid convergence γ. The grid convergence is used to convert the geodetic azimuth and grid azimuth, and vice versa. It can be calculated given either the geodetic coordinates (L, B) or grid coordinates (x, y). The derivations of their formulae are provided below.

To Compute γ Given Geodetic Coordinates (L, B)

From Fig. 6.17, on the projected parallel P', according to the geometric meaning of first-order derivative, we get:

$$\tan\gamma = \frac{\mathrm{d}x}{\mathrm{d}y}.$$

Taking the total differential of $x = f_1(q, l)$, $y = f_2(q, l)$ yields:

$$\mathrm{d}x = \frac{\partial x}{\partial q}\mathrm{d}q + \frac{\partial x}{\partial l}\mathrm{d}l,$$
$$\mathrm{d}y = \frac{\partial y}{\partial q}\mathrm{d}q + \frac{\partial y}{\partial l}\mathrm{d}l.$$

On the parallel curve $P'E'$, B is constant, and therefore in this case $\mathrm{d}q = 0$. Hence, we have:

$$\left.\begin{aligned}\mathrm{d}x &= \frac{\partial x}{\partial l}\mathrm{d}l \\ \mathrm{d}y &= \frac{\partial y}{\partial l}\mathrm{d}l\end{aligned}\right\},$$

so

$$\tan\gamma = \frac{\dfrac{\partial x}{\partial l}}{\dfrac{\partial y}{\partial l}}. \tag{6.77}$$

From (6.63) we have obtained the partial derivatives of x and y with respect to l. Inserting into the above equation and rearranging produces:

$$\tan\gamma = \sin B \cdot l + \frac{1}{3}\sin B\cos^2 B\left(1 + t^2 + 3\eta^2 + 2\eta^4\right)l^3 + \frac{1}{15}\sin B\cos^4 B\left(2 + 4t^2 + 2t^4\right)l^5.$$

With $\tan\gamma = x$
and $\gamma = \tan^{-1}x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 + \cdots = \tan\gamma - \frac{1}{3}\tan^3\gamma + \frac{1}{5}\tan^5\gamma + \cdots$
the result is:

$$\gamma'' = l''\sin B\left[1 + \frac{l''^2\cos^2 B}{3\rho''^2}\left(1 + 3\eta^2 + 2\eta^4\right) + \frac{l''^4\cos^4 B}{15\rho''^4}\left(2 - t^2\right)\right]. \tag{6.78}$$

Equation (6.78) is the formula for computing the grid convergence γ from geodetic coordinates (L, B). It can thus be seen that:

1. When $l = 0$, $\gamma = 0$ and when $B = 0$, $\gamma = 0$, i.e., the grid convergence is zero on both the central meridian and the equator.
2. Grid convergence γ is the odd function of l. l and γ are both considered to be positive when point P is east of the central meridian, and negative when P is west of the central meridian.
3. When the latitude B is constant, the value of γ increases with the increasing difference in longitude between the central meridian and point P.
4. When l is constant, the value of γ increases as latitude increases towards the poles. γ becomes greatest at the poles.

Equation (6.78) can be accurate to $0.001''$ when $l \leq 3.5^{\circ}$. wWhen $l \leq 2\ ^{\circ}$, the term that contains l^5 is less than $0.001''$ and can be neglected.

To Compute γ Given Plane Coordinates (x, y)

The formula for computing the grid convergence from given plane coordinates x, y can be obtained by making changes to (6.78), in which we replace l by Cartesian coordinates and B by B_f. We will derive the formula to the term that contains y^3.

B is replaced by B_f through expanding sin$B$ using the Taylor series:

$$\sin B = \sin\left[B_f - \left(B_f - B\right)\right] = \sin B_f - \cos B_f\left(B_f - B\right) - \cdots,$$

where $(B_f - B)$ is obtained by taking the principal terms in the first expression of (6.46):

$$\left(B_f - B\right) = \frac{t_f}{2M_f N_f}y^2 = \frac{t_f}{2N_f^2}y^2\left(1 + \eta_f^2\right).$$

Substituting gives:

$$\sin B = \cos B_f \left[t_f - \frac{t_f}{2N_f^2} y^2 \left(1 + \eta_f^2\right) - \cdots \right]. \tag{6.79}$$

In the same way, we get:

$$\begin{aligned} \cos B &= \cos\left[B_f - \left(B_f - B\right)\right] = \cos B_f + \sin B_f \left(B_f - B\right) - \cdots \\ &= \cos B_f + \sin B_f \frac{t_f}{2M_f N_f} y^2 = \cos B_f \left[1 + \frac{{t_f}^2}{2M_f N_f} y^2 \right]. \end{aligned} \tag{6.80}$$

We replace l with Cartesian coordinates, and apply the second expression in (6.46) to get:

$$\left.\begin{aligned} l'' \cos B_f &= \frac{\rho''}{N_f} y - \frac{\rho''}{6N_f^3} y^3 \left(1 + 2t_f^2 + \eta_f^2\right) + \cdots \\ l'' \cos {}^2 B_f &= \frac{\rho''^2}{N_f^2} y^2 - \cdots \end{aligned}\right\}. \tag{6.81}$$

Inserting (6.79), (6.80), and (6.81) into (6.78), and neglecting the terms containing $l^3\eta^4$ or higher powers yields:

$$\begin{aligned} \gamma'' &= \left[\frac{\rho''}{N_f} y - \frac{\rho''}{6N_f^3} y^3 \left(1 + 2t_f^2 + \eta_f^2\right) \right] \cdot \left[t_f - \frac{t_f}{2N_f^2} y^2 \left(1 + \eta_f^2\right) \right] \cdot \left[1 + \frac{1}{3{N_f}^2} y^2 \left(1 + 3\eta_f^2\right) \right] \\ &= \frac{\rho'' y}{N_f} t_f - \frac{\rho'' y^3}{6N_f^3} t_f \left(1 + 2t_f^2 + \eta_f^2\right) - \frac{\rho'' y^3}{2N_f^3} t_f \left(1 + \eta_f^2\right) + \frac{\rho'' y^3}{3N_f^3} t_f \left(1 + 3\eta_f^2\right). \end{aligned}$$

Finally one obtains:

$$\gamma'' = \frac{\rho'' y}{N_f} t_f - \frac{\rho'' y^3}{3N_f^3} t_f \left(1 + t_f^2 - \eta_f^2\right), \tag{6.82}$$

or, if calculated to terms containing y^5, then:

$$\gamma'' = \frac{\rho'' y}{N_f} t_f - \frac{\rho'' y^3}{3N_f^3} t_f \left(1 + t_f^2 - \eta_f^2\right) + \frac{\rho'' y^5}{15N_f^5} t_f \left(2 + 5t_f^2 + 3t_f^4\right). \tag{6.83}$$

γ can achieve an accuracy of $1''$ if calculated from (6.82) and $0.001''$ from (6.83). The subscript f indicates that the result is obtained from the footprint latitude, which is similar to the meaning it carries in the formula for the inverse solution of the Gauss projection.

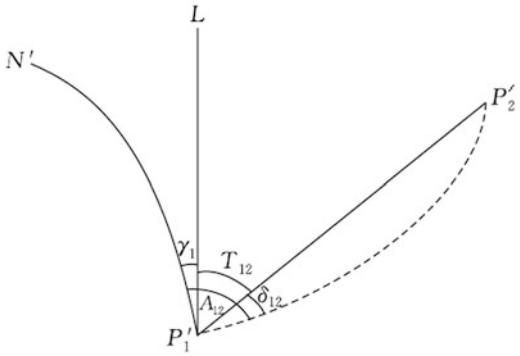


Fig. 6.18 Computation of grid bearing $T_{1.2}$. The *dashed line* indicates the projected curve

Practical Formula

Equations (6.78) and (6.83) can both be used as practical formulae for computer programming to achieve an accuracy of 0.001″. The value of B_f needed to calculate (6.83) can be obtained through iteration or from the direct formula based on $x = X$. To reach an accuracy of 0.0001″ we can expand the series in (6.83), and the result is as follows:

$$\gamma'' = \frac{\rho'' y}{N_f} t_f - \frac{\rho'' y^3}{3N_f^3} t_f \left(1 + t_f^2 - 5\eta_f^2\right) + \frac{\rho'' y^5}{15N_f^5} t_f \left(2 + 5t_f^2 + 3t_f^4 + 2\eta_f^2 + \eta_f^2 t_f^2\right).$$

For instance, given $L = 113°50'26.268''$ and $B = 31°33'22.293''$, according to (6.78) one gets $\gamma = +1°29'14.992''$. Again, given $x = 3{,}496{,}205.167$ m and $y = 269{,}759.797$ m, one obtains from (6.83) that $\gamma = +1°29'14.992''$.

6.5.5 *Computation of Grid Bearing*

As shown in Fig. 6.18, the angle between the curve $P_1'N'$ and the straight line $P_1'L$ is the grid convergence γ_1 of point P_1'. The angular difference between $P_1'L$ and the chord $\overline{P_1'P_2'}$ is the grid azimuth $T_{1.2}$ of P_1' in the direction of $P_1'P_2'$. As the Gauss projection is conformal, the angle between $P_1'N'$ and the projected curve $\overset{\frown}{P_1'P_2'}$ (dashed line in Fig. 6.18) on the plane is equivalent to the angle between P_1N and P_1P_2 on the ellipsoid, which is the geodetic azimuth $A_{1.2}$. From Fig. 6.18, we have

$$T_{1.2} = A_{1.2} - \gamma_1 - |\delta_{1.2}|.$$

The sign of the arc-to-chord correction $\delta_{1.2}$ in Fig. 6.18 is negative. $T_{1.2}$ can therefore be calculated from $A_{1.2}$ by applying the formula:

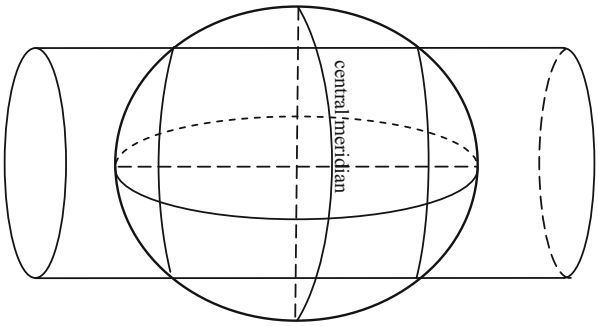


Fig. 6.19 Geometric description of the UTM projection

$$T_{1.2} = A_{1.2} - \gamma_1 + \delta_{1.2} \tag{6.84}$$

When converting the triangulation network on the ellipsoid to the Gauss plane, the grid azimuth on the initial side should be calculated according to (6.84). The geodetic azimuth $A_{1.2}$, according to Laplace's azimuth formula (5.67), can be obtained through the actually measured astronomical azimuth $\alpha_{1.2}$, namely:

$$A_{1.2} = \alpha_{1.2} - (\lambda_1 - L_1)\sin\varphi_1.$$

Combining the above two equations, we get:

$$T_{1.2} = \alpha_{1.2} - (\lambda_1 - L_1)\sin\varphi_1 - \gamma_1 + \delta_{1.2}, \tag{6.85}$$

which is the formula for computing the grid azimuth from the astronomical azimuth. In (6.85), λ_1 and L_1 are the astronomical longitude and geodetic longitude of P_1, and γ_1 is the grid convergence of P_1.

6.6 Universal Transverse Mercator Projection

6.6.1 Definition of UTM Projection

The Gauss projection is also called Transverse Mercator (TM) projection. Geometrically, it can be approximately perceived to be a transverse cylindrical conformal projection.

The Universal Transverse Mercator (UTM) projection can be understood geometrically as an equiangular transverse secant cylindrical projection. The UTM system divides the surface of the Earth (considered as a sphere) between 80°S latitude and 84°N latitude, as shown in Fig. 6.19. The cylindrical projection intersects the Earth at two lines (approximate to two meridians). No distortion occurs anywhere that the projection surface intersects with the Earth's surface. The scale factor along the central meridian is less than 1.

Conditions for the UTM projection are:

1. The projection is conformal
2. The central meridian is projected as a straight line
3. The central meridian and all distances have a scale factor of 0.9996 after projection

The scale factor along the central meridian is 0.9996 rather than 1, which is a reduction of 0.0004. For distances that are not close to the central meridian, the UTM has an advantage over the Gauss projection in reducing the amount of distance distortion and satisfies the needs of topographic maps.

Analogous to the Gauss projection, the UTM system divides the regions north of 84°N and south of 80°S into 60 longitudinal zones of 6°. These zones are numbered 1 through 60, starting at longitude 180°, and proceeding eastward (Beijing is in zone 50). To avoid negative coordinates for positions located west of the central meridian, the central meridian has been given a false y-coordinate of 500,000 m. A false x-coordinate of 10,000,000 m is allocated to the equator in the southern hemisphere.

6.6.2 Computational Formula for UTM Projection

UTM and Gauss projections are essentially the same. The previous two conditions for the UTM projection correspond to those for Gauss projection. The only difference between them lies in that for the UTM projection, the central meridian has a scale factor of 0.9996, rather than 1.0. It can be seen that UTM and Gauss projections are related by a similarity transformation, based on which, one can write out the formula for computing the UTM projection according to relevant formulae for the Gauss projection:

1. For the direct solution of the UTM projection, one can reckon the Gauss plane coordinates $(x_{\text{Gauss}}, y_{\text{Gauss}})$ using the formulae for the direct solution of the Gauss projection and obtain the UTM plane coordinates $(x_{\text{UTM}}, y_{\text{UTM}})$ according to the formulae below:

$$\begin{aligned} x_{\text{UTM}} &= 0.9996 x_{\text{Gauss}} \\ y_{\text{UTM}} &= 0.9996 y_{\text{Gauss}}. \end{aligned} \tag{6.86}$$

2. For the inverse solution of the UTM projection, we can compute the corresponding $(x_{\text{Gauss}}, y_{\text{Gauss}})$ given the $(x_{\text{UTM}}, y_{\text{UTM}})$ using the formulae below:

$$\begin{aligned} x_{\text{Gauss}} &= x_{\text{UTM}}/0.9996 \\ y_{\text{Gauss}} &= y_{\text{UTM}}/0.9996. \end{aligned} \tag{6.87}$$

Then, apply the formulae for the inverse solution of the Gauss projection to calculate (L, B).

Table 6.6 Numerical values of the distance distortion for different longitudes (l) and latitudes (B)

B	$l = 0°$	$l = 1°$	$l = 2°$	$l = 3°$
90°	−0.00040	−0.00040	−0.00040	−0.00040
80°	−0.00040	−0.00040	−0.00038	−0.00036
70°	−0.00040	−0.00038	−0.00033	−0.00024
60°	−0.00040	−0.00036	−0.00025	−0.00006
50°	−0.00040	−0.00034	−0.00015	+0.00017
40°	−0.00040	−0.00031	−0.00004	+0.00041
30°	−0.00040	−0.00028	+0.00006	+0.00063
20°	−0.00040	−0.00027	+0.00014	+0.00081
10°	−0.00040	−0.00026	+0.00019	+0.00094
0°	−0.00040	−0.00025	+0.00021	+0.00098

3. The formula for arc-to-chord correction of the UTM projection agrees with that of Gauss projection. Conversions like (6.87) are not needed because the arc-to-chord correction is the correction applied to compensate for distortion of straight lines when projected, which is fundamentally independent of projection method.
4. The relationship between the scale factor m_{UTM} of the UTM projection and m_{Gauss} of the Gauss projection is:

$$m_{\text{UTM}} = 0.9996 m_{\text{Gauss}}. \tag{6.88}$$

5. To compute the distance correction of the UTM projection, first we transform the coordinates according to (6.87) and calculate the Gauss plane distance D_{Gauss} according to the formula for distance corrections in the Gauss projection; then, we reckon the UTM plane distance based on the formula

$$D_{\text{UTM}} = 0.9996 D_{\text{Gauss}}. \tag{6.89}$$

6. The formula for meridian convergence on the UTM plane completely corresponds to that on the Gauss plane.

The distance distortion of the UTM projection can be analyzed using (6.88). Table 6.6 presents the distortion of distances given the different longitudes and latitudes.

Table 6.6 shows that the distortion of distance along the central meridian is −0.00040, namely, the central meridian has a scale factor of 0.9996. This allows the greatest distortion of distance at $B = 0^{\text{o}}$, $l = 3^{\text{o}}$ to be smaller than 0.001. The two secant lines (about ±180 km away from the central meridian on the equator, i.e., at about $\pm 1°40'$) have no distortion of distances. Distances become more distorted away from these two secant lines. The distance distortion will have a negative value inside the two secants and a positive value outside.

Review and Study Questions

1. Why are geodetic elements on the ellipsoid projected onto a plane?
2. What is the conformality of conformal projection? Please demonstrate.
3. Write out the Cauchy–Riemann differential equation for conformal projection from the ellipsoid onto the plane and from the plane onto the ellipsoid, and give the relationship between isometric latitude and geodetic latitude.
4. What is zone division and how can an area of interest be divided into zones?
5. What are the three conditions for the Gauss projection and how does one express these conditions with formulae?
6. Draw a geometric figure to explain the expansion point of the direct and inverse solutions of Gauss projection, and explain the term "footprint latitude."
7. Illustrate the basics of the transformation between adjacent zones and how to transform coordinates between 3° zones and 6° zones.
8. What is involved in the reduction of a horizontal geodetic control network to the Gauss plane?
9. Derive the approximate formula for arc-to-chord correction with an accuracy of 0.1 m.
10. Derive the formula for checking arc-to-chord correction.
11. Analyze the distortion property of the Gauss projection according to the formula for distance distortion $m - 1 = \frac{y^2}{2R^2} + \frac{y^4}{24R^4}$.
12. Analyze the signs of the two corrected values when reducing the terrestrial observed distance to the plane.
13. Analyze the desired accuracy for coordinates in computations of distance correction and arc-to-chord correction of the first-order triangulateration control network (let $R = 6{,}400$ km, y and $\Delta x = 300$ km, and $S = 50$ km).
14. Draw a geometric figure to derive the relational expression for computing the grid convergence:

$$\tan\gamma = \frac{\dfrac{\partial x}{\partial l}}{\dfrac{\partial y}{\partial l}}.$$

15. Explain the concepts of the Universal Transverse Mercator (UTM) projection.

Chapter 7
Establishment of Geodetic Coordinate Systems

To describe the state of an object, we have to make clear what it is referenced to. In the context of geodetic surveying, other than choosing a reference, one still needs to carry out spatial positioning and orientation and specify the unit of measurement (such as time scale, spatial scale, etc.). Therefore, it is necessary to establish a terrestrial reference coordinate system (also known as reference system or coordinate system, which are interpreted as synonyms here). Mathematically, it is unreasonable to judge the merits and demerits of a coordinate system. Nevertheless, from a physical and functional perspective, we should choose the proper reference system, taking into account the operability and convenience of the issues being studied.

This chapter discusses the principles for establishing classical and modern geodetic coordinate systems, establishes the transformation models between different coordinate systems, and provides an overview of the geodetic coordinate systems in China and throughout the world.

7.1 Euler Angles in Geodetic Coordinate Systems

7.1.1 Vector Analysis in Coordinate Transformations

In Fig. 7.1, two spatial Cartesian coordinate systems are introduced, O-XYZ and O-$X'Y'Z$. Our discussion involves only the coordinate transformations under rotation, so their origins are assumed to be coincident. The direction angles of the coordinate axes OX', OY', and OZ' of O-$X'Y'Z'$ with respect to the axes OX, OY, OZ of the coordinate system O-XYZ are $\alpha_1, \beta_1, \gamma_1$; $\alpha_2, \beta_2, \gamma_2$; and $\alpha_3, \beta_3, \gamma_3$, respectively.

Let $\vec{r}$ denote the radius vector of a point M in space in relation to the coordinate system O-XYZ and $\vec{r}'$ represent the radius vector of the same point in space relative to the system O - $X'Y'Z'$; then it is obvious that:

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5_7,

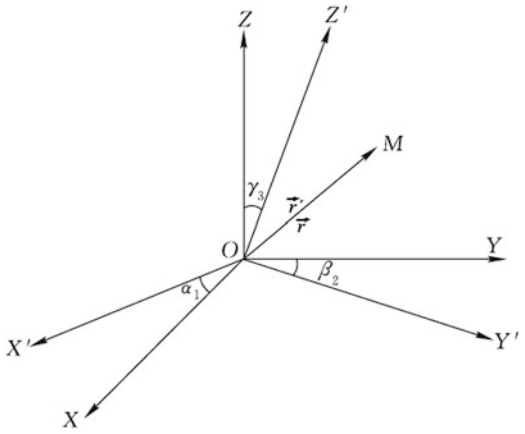


Fig. 7.1 Direction angles

$$\vec{r}' = \vec{r}$$

By applying the expression of their components, the above equation becomes:

$$X'\vec{i}' + Y'\vec{j}' + Z'\vec{k}' = X\vec{i} + Y\vec{j} + Z\vec{k}, \tag{7.1}$$

where $\vec{i}'$, $\vec{j}'$, and $\vec{k}'$ are the basic unit vectors of O - $X'Y'Z'$, while $\vec{i}$, $\vec{j}$, and $\vec{k}$ are the basic unit vectors of O - XYZ; X', Y', Z' and X, Y, Z are the components of $\vec{r}$ and $\vec{r}'$, respectively.

Taking the dot product (scalar product) of both sides of (7.1) with $\vec{i}'$ and then with $\vec{j}'$ and $\vec{k}'$, respectively, yields:

$$\begin{aligned} X' &= X\vec{i}' \cdot \vec{i} + Y\vec{i}' \cdot \vec{j} + Z\vec{i}' \cdot \vec{k}, \\ Y' &= X\vec{j}' \cdot \vec{i} + Y\vec{j}' \cdot \vec{j} + Z\vec{j}' \cdot \vec{k}, \end{aligned}$$

and

$$Z' = X\vec{k}' \cdot \vec{i} + Y\vec{k}' \cdot \vec{j} + Z\vec{k}' \cdot \vec{k}.$$

From the definition of the dot product of two vectors, we have:

$$\begin{aligned} \vec{i}' \cdot \vec{i} &= \cos\left(\vec{i}', \vec{i}\right) = \cos\alpha_1, \\ \vec{i}' \cdot \vec{j} &= \cos\left(\vec{i}', \vec{j}\right) = \cos\beta_1, \\ \vec{i}' \cdot \vec{k} &= \cos\left(\vec{i}', \vec{k}\right) = \cos\gamma_1, \\ &\vdots \\ \vec{k}' \cdot \vec{k} &= \cos\left(\vec{k}', \vec{k}\right) = \cos\gamma_3. \end{aligned}$$

Hence, the above equation can be rewritten as:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} \cos\alpha_1 & \cos\beta_1 & \cos\gamma_1 \\ \cos\alpha_2 & \cos\beta_2 & \cos\gamma_2 \\ \cos\alpha_3 & \cos\beta_3 & \cos\gamma_3 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \tag{7.2}$$

where the coefficients are termed the transformation coefficients. The first, second, and third row of the coefficient matrix are the respective coordinates of $\vec{i}', \vec{j}'$, and $\vec{k}'$ in $O\text{-}XYZ$, and $\vec{i}' \cdot \vec{i}' = 1$, $\vec{j}' \cdot \vec{j}' = 1$, $\vec{k}' \cdot \vec{k}' = 1$, $\vec{i}' \cdot \vec{j}' = 0$, $\vec{i}' \cdot \vec{k}' = 0$, and $\vec{j}' \cdot \vec{k}' = 0$. Hence, the nine direction angles in (7.2) should satisfy the six relational expressions below:

$$\left.\begin{array}{l} \cos^2\alpha_1 + \cos^2\beta_1 + \cos^2\gamma_1 = 1 \\ \cos^2\alpha_2 + \cos^2\beta_2 + \cos^2\gamma_2 = 1 \\ \cos^2\alpha_3 + \cos^2\beta_3 + \cos^2\gamma_3 = 1 \\ \cos\alpha_1\cos\alpha_2 + \cos\beta_1\cos\beta_2 + \cos\gamma_1\cos\gamma_2 = 0 \\ \cos\alpha_2\cos\alpha_3 + \cos\beta_2\cos\beta_3 + \cos\gamma_2\cos\gamma_3 = 0 \\ \cos\alpha_3\cos\alpha_1 + \cos\beta_3\cos\beta_1 + \cos\gamma_3\cos\gamma_1 = 0 \end{array}\right\}. \tag{7.3}$$

Theoretically, only three direction angles out of the nine are independent, meaning that one can employ any three arbitrary independent direction angles to represent the remaining six. When studying the positioning and orientation of the ellipsoid as well as the transformation between different coordinate systems, we are more concerned with the angles between the corresponding coordinate axes. Therefore, we have chosen α_1 (the angle between the two X-axes), β_2 (the angle between the two Y-axes), and γ_3 (the angle between the two Z-axes). Of these, β_2 and γ_3 are our top concerns. Since the Z and Z' axes coincide with their respective minor axis of the ellipsoid, γ_3 then denotes the angle between the minor axes of two ellipsoids that are non-parallel to each other. ZOX and $Z'OX'$ are their respective planes of initial geodetic meridian. Hence, β_2 represents the angle between two planes of initial geodetic meridian that are not parallel to each other.

7.1.2 *Coordinate Transformations in Terms of Euler Angles as Rotation Parameters*

The formulae are long-winded if α_1, β_2, and γ_3 are chosen as the independent direction angles. Thus, we opt for another three mutually independent parameters to represent the direction angles. These three parameters are the angles produced by three successive rotations about the coordinate axis, defined as the Euler angles (see e.g., Grewal et al. 2001). Euler angles are different from the angles between the corresponding axes of two Cartesian coordinate systems, but an analytical relationship can be established between them. The Euler angles in the geodetic coordinate system are also referred to as rotation parameters.

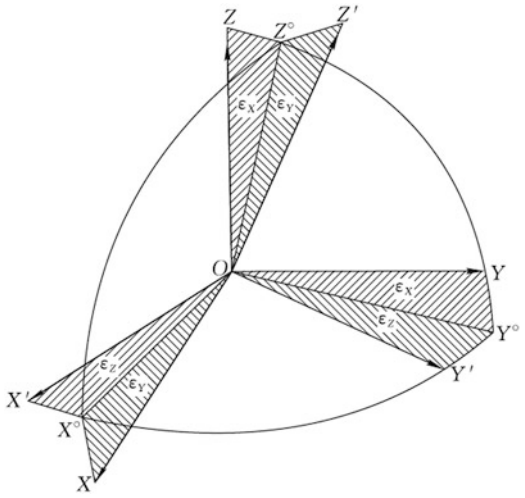


Fig. 7.2 Euler angles

Euler angles are usually described as shown in Fig. 7.2. Assign ε_X, ε_Y, and ε_Z as Euler angles, and the rotations of the coordinate system are as follows:

First, rotate OX' to OX° and correspondingly OY' to OY° around the OZ' axis by the rotation angle of ε_Z. Then rotate OZ' to OZ° around the OY° axis, and correspondingly OX° to OX, and the rotation angle is ε_Y. Finally, rotate OZ° to axis OZ around the OX axis; OY° is correspondingly rotated to OY, and the angle of rotation is ε_X.

Hence, we can write:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{R_X}(\varepsilon_X)\mathbf{R_Y}(\varepsilon_Y)\mathbf{R_Z}(\varepsilon_Z)\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}, \tag{7.4}$$

where $\mathbf{R_X}(\varepsilon_X)$, $\mathbf{R_Y}(\varepsilon_Y)$, and $\mathbf{R_Z}(\varepsilon_Z)$ are the rotation matrices, and the expressions are given by:

$$\left.\begin{aligned} \mathbf{R_X}(\varepsilon_X) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varepsilon_X & \sin\varepsilon_X \\ 0 & -\sin\varepsilon_X & \cos\varepsilon_X \end{bmatrix} \\ \mathbf{R_Y}(\varepsilon_Y) &= \begin{bmatrix} \cos\varepsilon_Y & 0 & -\sin\varepsilon_Y \\ 0 & 1 & 0 \\ \sin\varepsilon_Y & 0 & \cos\varepsilon_Y \end{bmatrix} \\ \mathbf{R_Z}(\varepsilon_Z) &= \begin{bmatrix} \cos\varepsilon_Z & \sin\varepsilon_Z & 0 \\ -\sin\varepsilon_Z & \cos\varepsilon_Z & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}\right\}. \tag{7.5}$$

Substituting (7.5) into (7.4) yields:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \cos\varepsilon_Y\cos\varepsilon_Z & \cos\varepsilon_Y\sin\varepsilon_Z & -\sin\varepsilon_Y \\ -\cos\varepsilon_X\sin\varepsilon_Z+\sin\varepsilon_X\sin\varepsilon_Y\cos\varepsilon_Z & \cos\varepsilon_X\cos\varepsilon_Z+\sin\varepsilon_X\sin\varepsilon_Y\sin\varepsilon_Z & \sin\varepsilon_X\cos\varepsilon_Y \\ \sin\varepsilon_X\sin\varepsilon_Z+\cos\varepsilon_X\sin\varepsilon_Y\cos\varepsilon_Z & -\sin\varepsilon_X\cos\varepsilon_Z+\cos\varepsilon_X\sin\varepsilon_Y\sin\varepsilon_Z & \cos\varepsilon_X\cos\varepsilon_Y \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} \tag{7.6}$$

When ε_X, ε_Y, and ε_Z are very small, we neglect the second-order small quantities to obtain:

$$\left.\begin{array}{l} \cos\varepsilon_X \approx \cos\varepsilon_Y \approx \cos\varepsilon_Z \approx 1 \\ \sin\varepsilon_X \approx \varepsilon_X,\ \sin\varepsilon_Y \approx \varepsilon_Y,\ \sin\varepsilon_Z \approx \varepsilon_Z \\ \sin\varepsilon_X\sin\varepsilon_Y \approx \sin\varepsilon_Y\sin\varepsilon_Z \approx \sin\varepsilon_Z\sin\varepsilon_X \approx 0 \end{array}\right\}.$$

Then (7.6) can be written as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & \varepsilon_Z & -\varepsilon_Y \\ -\varepsilon_Z & 1 & \varepsilon_X \\ \varepsilon_Y & -\varepsilon_X & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}, \tag{7.7}$$

where the coefficient matrix is also called the differential rotation matrix. Comparing (7.2) and (7.6) gives:

$$\left.\begin{array}{l} \cos\gamma_3 = \cos\varepsilon_X\cos\varepsilon_Y \\ \cos\beta_2 = \cos\varepsilon_X\cos\varepsilon_Z + \sin\varepsilon_X\sin\varepsilon_Y\sin\varepsilon_Z \\ \cos\alpha_1 = \cos\varepsilon_Y\cos\varepsilon_Z \end{array}\right\}. \tag{7.8}$$

Omitting the small terms higher than third order created by mutual multiplication of ε_X, ε_Y, and ε_Z in (7.8) produces:

$$\left.\begin{array}{l} \gamma_3 \approx \sqrt{\varepsilon_X^2+\varepsilon_Y^2} \\ \beta_2 \approx \sqrt{\varepsilon_X^2+\varepsilon_Z^2} \\ \alpha_1 \approx \sqrt{\varepsilon_Y^2+\varepsilon_Z^2} \end{array}\right\}. \tag{7.9}$$

Equation (7.7) shows that by neglecting the small quantities of the second-order, the rotation matrices are commutative.

The three rotation angles ε_Z, ε_Y, and ε_X are called the yaw, pitch, and roll, respectively, in describing the vehicle's attitude and are used to represent the precession, rotation (spin), and nutation, respectively while studying the rotation of rigid bodies.

7.1.3 Generalized Formulae for Deflection of the Vertical and Laplace Azimuth

In Sect. 5.5, on condition that the minor axis of the ellipsoid is parallel to the Earth's rotational axis and the planes of the initial geodetic and astronomical meridians are parallel to each other, we derived the formulae for the deflection of the vertical and the Laplace azimuth, as follows:

$$\left.\begin{aligned}\xi &= \varphi - B \\ \eta &= (\lambda - L)\cos\varphi\end{aligned}\right\},$$

and

$$A = \alpha - (\lambda - L)\sin\varphi = \alpha - \eta\tan\varphi.$$

When the minor axis of the ellipsoid is non-parallel to the Earth's rotational axis at a certain epoch and the initial geodetic meridian plane is non-parallel to the initial astronomical meridian plane, Euler angles ε_X, ε_Y, and ε_Z arise. In this case, the corresponding correction terms should be applied to the above formulae.

Here, we provide the formulae for the deflection of the vertical and Laplace azimuth when Euler angles ε_X, ε_Y, and ε_Z exist (without derivation):

$$\left.\begin{aligned}\xi &= \varphi - B + \sin\lambda\varepsilon_X - \cos\lambda\varepsilon_Y \\ \eta &= (\lambda - L)\cos\varphi - \cos\lambda\sin\varphi\varepsilon_X - \sin\lambda\sin\varphi\varepsilon_Y + \cos\varphi\varepsilon_Z\end{aligned}\right\}, \tag{7.10}$$

$$A = \alpha - \eta\tan\varphi - (\varepsilon_Y\sin\lambda + \varepsilon_X\cos\lambda)\sec\varphi, \tag{7.11}$$

where (7.10) is the generalized formula for the deflection of the vertical and (7.11) is the that for the Laplace azimuth.

7.2 Transformation Between Different Geodetic Coordinate Systems

7.2.1 Transformation Between Different Geodetic Cartesian Coordinate Systems

Different geodetic coordinate systems arise because the data, models, parameters, and data processing methods employed in the establishment of geodetic coordinate system are different. In practice, as often as not, we have to solve the problem of unification between different coordinate systems.

In Fig. 7.3, O_{new}—$X_{new}\ Y_{new}\ Z_{new}$ and O_{old}—$X_{old}\ Y_{old}\ Z_{old}$ are two geodetic Cartesian coordinate systems (geodetic spatial rectangular coordinate systems). Their coordinate origins do not coincide, i.e., there are three translation parameters, $\Delta X_0, \Delta Y_0, \Delta Z_0$, denoting the components of the old coordinate origin with respect to the new coordinate origin along the three coordinate axes. The coordinate axes of the two systems are mutually non-parallel, generally less than $1''$ for classical geodetic coordinate systems, meaning that Euler angles $\varepsilon_x, \varepsilon_y, \varepsilon_z$ exist (known also as the three rotation parameters). Apparently, these two systems can be made coincident under translation and rotation. According to (7.7), we get:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix} + \begin{bmatrix} 1 & \varepsilon_Z & -\varepsilon_Y \\ -\varepsilon_Z & 1 & \varepsilon_X \\ \varepsilon_Y & -\varepsilon_X & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old}. \tag{7.12}$$

For various reasons, there will also be a difference in scale while establishing the two systems. Assume that S_{new} and S_{old} are the measurements of the same distance in space in the new and old coordinate systems; then we can define

$$\Delta m = \frac{S_{new} - S_{old}}{S_{old}} = m - 1$$

as the scale factor of the two coordinate systems. Here, Δm is homogeneous and is independent of the point position and direction. Hence, the old coordinates can be improved in accordance with the scale of the new coordinate system, as $X = X_{old} + \Delta m X_{old}$, $Y = Y_{old} + \Delta m Y_{old}$ and $Z = Z_{old} + \Delta m Z_{old}$.

In (7.12), considering the effect of the scale factor means improving the $(X, Y, Z)^T_{old}$ in accordance with the above relations, namely:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix} + \begin{bmatrix} 1 & \varepsilon_Z & -\varepsilon_Y \\ -\varepsilon_Z & 1 & \varepsilon_X \\ \varepsilon_Y & -\varepsilon_X & 1 \end{bmatrix} \begin{bmatrix} X_{old} + \Delta m X_{old} \\ Y_{old} + \Delta m Y_{old} \\ Z_{old} + \Delta m Z_{old} \end{bmatrix}.$$

Disregarding the second-order small quantities and rearranging gives:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = (1 + \Delta m) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old} + \begin{bmatrix} 0 & \varepsilon_Z & -\varepsilon_Y \\ -\varepsilon_Z & 0 & \varepsilon_X \\ \varepsilon_Y & -\varepsilon_X & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old} + \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix}, \tag{7.13}$$

which is called the Bursa–Wolf transformation model (or, simply, Bursa model; see Thomson 1976) with seven transformation parameters $\Delta X_0, \Delta Y_0, \Delta Z_0, \varepsilon_X, \varepsilon_Y, \varepsilon_Z$, and Δm. The linear equation in terms of these parameters is:

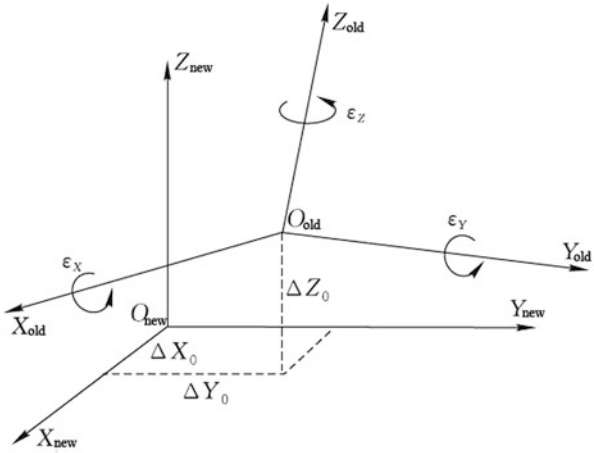


Fig. 7.3 Transformations between geodetic Cartesian coordinate systems

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} - \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old} = \begin{bmatrix} 1 & 0 & 0 & 0 & -Z_{old} & Y_{old} & X_{old} \\ 0 & 1 & 0 & Z_{old} & 0 & -X_{old} & Y_{old} \\ 0 & 0 & 1-Y_{old} & X_{old} & 0 & Z_{old} & \end{bmatrix} \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \\ \varepsilon_X \\ \varepsilon_Y \\ \varepsilon_Z \\ \Delta m \end{bmatrix}. \tag{7.14}$$

In (7.13), if $\varepsilon_X = \varepsilon_Y = \varepsilon_Z = 0$ and $\Delta m = 0$, then it is called the three-parameter formula, indicating that the scales of the two geodetic Cartesian coordinate systems are consistent and the corresponding coordinate axes are mutually parallel. Likewise, in (7.13), by leaving out certain parameters, we can obtain the four-parameter, five-parameter, and six-parameter transformation formulae.

In order to obtain the seven transformation parameters in (7.14), at least three points with two sets of coordinates, both the old and the new (known as common points) are needed. The transformation parameters will be solved according to the principle of adjustment.

Actually, the accuracy of the common point coordinates and other factors such as the number and geometric distribution of the common points all exert an influence on the solution of transformation parameters. Thus, in practical cases, we have to choose a certain number of common points with relatively high accuracy and with even distribution and wide coverage.

After carrying out adjustment computations of the difference between the new and old coordinates as the observed quantity in (7.14), we will obtain the correction to the observed quantity. This shows that the new coordinates transformed from the old coordinates of the common points according to (7.14) are not completely

equivalent to the values of the given new coordinates. In practice, however, the coordinate values of the given points are often required to be constant and unchanging. In order to settle this issue, the transformed value of the common points can be corrected to the known values and one can deal with the transformed values of the non-common points, e.g., calculate the correction to the transformed values of the non-common points using the weighted average according to the formula below:

$$V' = \frac{\sum_{1}^{n} p_i v_i}{\sum_{1}^{n} p_i},$$

where n denotes the number of the common points. The weight of the ith common point can be defined by the distance (S_i) between the non-common point and the common point. We can set $p_i = \frac{1}{S_i^2}$, v_i is the correction to the coordinate value of the ith common point, namely $v_i = given\ value - transformed\ value$, and the coordinate of the common point adopts the given value. This is only one method for an interpolation of the residuals, which is not a similar transformation since the transformed network of identical points might lose its shape.

7.2.2 Transformation Between Different Geodetic Coordinate Systems

As indicated above, the transformation formulae for different geodetic Cartesian coordinate systems generally involve seven parameters: three translations, three rotations, and one scaling. For the transformation between different geodetic coordinate systems, two additional transformation parameters are needed, namely the different Earth ellipsoid parameters corresponding to the two types of geodetic coordinate systems. The transformation formula for different geodetic coordinate systems is also referred to as the geodetic coordinate differential formula or ellipsoid transformation differential formula. When inclusive of the rotation and scale parameters, it is called the generalized differential formula for geodetic coordinates or generalized differential formula for ellipsoid transformation.

Given that the relationship between the geodetic Cartesian coordinates and the geodetic coordinates of a given point in space is:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (N+H)\cos B \cos L \\ (N+H)\cos B \sin L \\ \left[N(1-e^2)+H\right]\sin B \end{bmatrix},$$

we can see that X, Y, and Z are the functions of L, B, H, a, and f (or e^2). When these variables are differentiated respectively as dL, dB, dH, da, and df, by taking the total differential of dX, dY, and dZ, one gets:

$$\begin{bmatrix} \mathrm{d}X \\ \mathrm{d}Y \\ \mathrm{d}Z \end{bmatrix} = \mathbf{J} \begin{bmatrix} \mathrm{d}L \\ \mathrm{d}B \\ \mathrm{d}H \end{bmatrix} + \mathbf{A} \begin{bmatrix} \mathrm{d}a \\ \mathrm{d}f \end{bmatrix}, \tag{7.15}$$

with

$$\begin{aligned} \mathbf{J} &= \begin{bmatrix} \dfrac{\partial X}{\partial L} & \dfrac{\partial X}{\partial B} & \dfrac{\partial X}{\partial H} \\ \dfrac{\partial Y}{\partial L} & \dfrac{\partial Y}{\partial B} & \dfrac{\partial Y}{\partial H} \\ \dfrac{\partial Z}{\partial L} & \dfrac{\partial Z}{\partial B} & \dfrac{\partial Z}{\partial H} \end{bmatrix} \\ &= \begin{bmatrix} -(N+H)\cos B \sin L & -(M+H)\sin B \cos L & \cos B \cos L \\ (N+H)\cos B \cos L & -(M+H)\sin B \sin L & \cos B \sin L \\ 0 & (M+H)\cos B & \sin B \end{bmatrix}, \end{aligned}$$

This is called a Jacobian matrix. For the solution to its inverse matrix, readers can refer to the literature (e.g., Zhu 1986).

In addition:

$$\mathbf{A} = \begin{bmatrix} \dfrac{\partial X}{\partial a} & \dfrac{\partial X}{\partial f} \\ \dfrac{\partial Y}{\partial a} & \dfrac{\partial Y}{\partial f} \\ \dfrac{\partial Z}{\partial a} & \dfrac{\partial Z}{\partial f} \end{bmatrix} = \begin{bmatrix} \dfrac{N}{a}\cos B \cos L & \dfrac{M}{1-f}\cos B \cos L \sin^2 B \\ \dfrac{N}{a}\cos B \sin L & \dfrac{M}{1-f}\cos B \sin L \sin^2 B \\ \dfrac{N}{a}\left(1-e^2\right)\sin B & -\dfrac{M}{1-f}\sin B\left(1+\cos^2 B - e^2 \sin^2 B\right) \end{bmatrix}.$$

It follows from (7.15) that:

$$\begin{bmatrix} \mathrm{d}L \\ \mathrm{d}B \\ \mathrm{d}H \end{bmatrix} = \mathbf{J}^{-1} \begin{bmatrix} \mathrm{d}X \\ \mathrm{d}Y \\ \mathrm{d}Z \end{bmatrix} - \mathbf{J}^{-1}\mathbf{A} \begin{bmatrix} \mathrm{d}a \\ \mathrm{d}f \end{bmatrix}, \tag{7.16}$$

where

$$\begin{bmatrix} \mathrm{d}X \\ \mathrm{d}Y \\ \mathrm{d}Z \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} - \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old},$$

and

$$\begin{bmatrix} \mathrm{d}L \\ \mathrm{d}B \\ \mathrm{d}H \end{bmatrix} = \begin{bmatrix} L \\ B \\ H \end{bmatrix}_{new} - \begin{bmatrix} L \\ B \\ H \end{bmatrix}_{old}.$$

In order to derive the inverse matrix $\mathbf{J}^{-1}$, we decompose $\mathbf{J}$ into the product of two matrices:

$$\mathbf{J} = \mathbf{SH}, \tag{7.17}$$

with

$$\mathbf{S} = \begin{bmatrix} -\sin L & -\sin B \cos L & \cos B \cos L \\ \cos L & -\sin B \sin L & \cos B \sin L \\ 0 & \cos B & \sin B \end{bmatrix},$$

and

$$\mathbf{H} = \begin{bmatrix} (N+H)\cos B & 0 & 0 \\ 0 & M+H & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

According to the invertible matrix theorem, we get:

$$\mathbf{J}^{-1} = \mathbf{H}^{-1}\mathbf{S}^{-1}. \tag{7.18}$$

$\mathbf{H}$ is the diagonal matrix and its inverse matrix is:

$$\mathbf{H}^{-1} = \begin{bmatrix} \dfrac{1}{(N+H)\cos B} & 0 & 0 \\ 0 & \dfrac{1}{M+H} & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7.19}$$

$\mathbf{S}$ is an orthogonal matrix; hence, we have:

$$\mathbf{S}^{-1} = \mathbf{S}^{\mathrm{T}} = \begin{bmatrix} -\sin L & \cos L & 0 \\ -\sin B \cos L & -\sin B \sin L & \cos B \\ \cos B \cos L & \cos B sibL & \sin B \end{bmatrix}. \tag{7.20}$$

Substituting (7.19) and (7.20) into (7.18) yields:

$$\mathbf{J}^{-1} = \begin{bmatrix} -\dfrac{\sin L}{(N+H)\cos B} & \dfrac{\cos L}{(N+H)\cos B} & 0 \\ -\dfrac{\sin B \cos L}{M+H} & -\dfrac{\sin B \sin L}{M+H} & \dfrac{\cos B}{M+H} \\ \cos B \cos L & \cos B \sin L & \sin B \end{bmatrix}. \tag{7.21}$$

We insert the Bursa–Wolf model (7.13) into (7.16) to get:

$$\begin{bmatrix} \mathrm{d}L \\ \mathrm{d}B \\ \mathrm{d}H \end{bmatrix} = \mathbf{J}^{-1} \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix} + \mathbf{J}^{-1}\mathbf{Q}\mathbf{B} + \mathbf{J}^{-1} m \mathbf{B} - \mathbf{J}^{-1}\mathbf{A} \begin{bmatrix} \mathrm{d}a \\ \mathrm{d}f \end{bmatrix}, \tag{7.22}$$

with

$$\mathbf{Q} = \begin{bmatrix} 0 & \varepsilon_Z & -\varepsilon_Y \\ -\varepsilon_Z & 0 & \varepsilon_X \\ \varepsilon_Y & -\varepsilon_X & 0 \end{bmatrix},$$

and

$$\mathbf{B} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{old} = \begin{bmatrix} (N+H)\cos B \cos L \\ (N+H)\cos B \sin L \\ \left[N(1-e^2)+H\right]\sin B \end{bmatrix}.$$

In the above equation, L, B, and H are the old geodetic coordinates and the subscripts "old" are all omitted. Rearranging the above equation gives:

$$
\begin{bmatrix} \mathrm{d}L \\ \mathrm{d}B \\ \mathrm{d}H \end{bmatrix} = \begin{bmatrix} -\dfrac{\sin L}{(N+H)\cos B}\rho'' & \dfrac{\cos L}{(N+H)\cos B}\rho'' & 0 \\ -\dfrac{\sin B\cos L}{M+H}\rho'' & -\dfrac{\sin B\sin L}{M+H}\rho'' & \dfrac{\cos B}{M+H}\rho'' \\ \cos B\cos L & \cos B\sin L & \sin B \end{bmatrix} \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix}
$$

$$
+ \begin{bmatrix} \tan B\cos L & \tan B\sin L & -1 \\ -\sin L & \cos L & 0 \\ -\dfrac{Ne^2\sin B\cos B\sin L}{\rho''} & \dfrac{Ne^2\sin B\cos B\cos L}{\rho''} & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_X'' \\ \varepsilon_Y'' \\ \varepsilon_Z'' \end{bmatrix}
$$

$$
+ \begin{bmatrix} 0 \\ -\dfrac{N}{M}e^2\sin B\cos B\rho'' \\ N(1-e^2\sin^2 B) \end{bmatrix} \Delta m
$$

$$
+ \begin{bmatrix} 0 & 0 \\ \dfrac{N}{(M+H)a}e^2\sin B\cos B\rho'' & \dfrac{M(2-e^2\sin^2 B)}{(M+H)(1-f)}\sin B\cos B\rho'' \\ -\dfrac{N}{a}\left(1-e^2\sin^2 B\right) & \dfrac{M}{1-f}\left(1-e^2\sin^2 B\right)\sin^2 B \end{bmatrix} \begin{bmatrix} \mathrm{d}a \\ \mathrm{d}f \end{bmatrix}, \tag{7.23}
$$

where $\mathrm{d}L$ and $\mathrm{d}B$ are measured in arcseconds. Equation (7.23) is the generalized differential formula for geodetic coordinates concerning the seven parameters and changes in the size of the ellipsoid. It should be noted from the formula that $\mathrm{d}a$, $\mathrm{d}f$, ΔZ_0, and Δm have no effect on the geodetic longitude (i.e., $\mathrm{d}L = 0$ here). The geodetic latitude and ellipsoidal height are independent of ε_Z. Disregarding the effect of the rotation and scale parameters, equation (7.23) is the general differential formula for geodetic coordinates.

7.2.3 *Grid Model of Coordinate Transformation*

Analogous to a grid model of deflections of the vertical and height anomalies, we can also establish the grid model of coordinate transformation. By making use of the coordinate values of the two geodetic coordinate systems on the common points

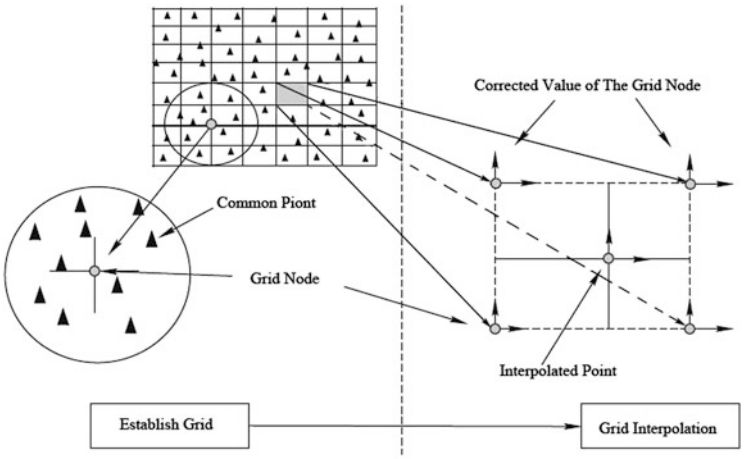


Fig. 7.4 Principle for grid coordinate transformation

and certain mathematical models (e.g., the least curvature principle, least squares collation, polynomial regression, and Bursa–Wolf model, etc.), we can compute the differences between the longitude and latitude coordinates of the grid nodes at certain distance intervals and establish the coordinate transformation grid model. Then, all we need to do is use the quantities of coordinate transformation of the four neighboring grid nodes of the points awaiting transformation to compute the quantities of coordinate transformation based on the bilinear interpolation formula (see Fig. 7.4). This approach is generally applied to the high accuracy transformation between the sheet lines of a topographic map and the square grids.

7.3 Classical Methods for Ellipsoid Orientation

7.3.1 Geodetic Origin Data and Ellipsoid Orientation

In classical geodesy, the ellipsoid orientation is meant to establish the geodetic coordinate system, i.e., to determine, under certain conditions, the position of the Earth ellipsoid with defined elements relative to the geoid, so as to obtain the reference surface and the geodetic origin data for geodetic computations.

Ellipsoid orientation means: (1) to determine the position of the center of an ellipsoid (abbreviated as positioning) and (2) to determine the direction of the coordinate axes of the Cartesian coordinate system with its origin at the center of the ellipsoid, i.e., to determine the pointing direction of the minor axis of an ellipsoid and the plane of the initial geodetic meridian (abbreviated as orientation).

The origin (initial point) from which the geodetic coordinates of points in the national horizontal geodetic control network are calculated is called the geodetic

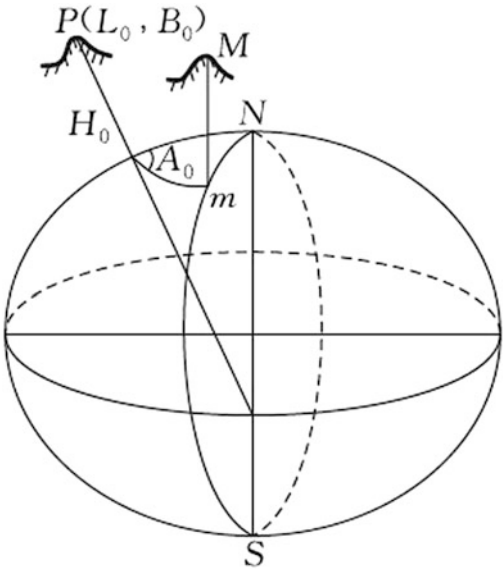


Fig. 7.5 Geodetic origin data and ellipsoid orientation

origin. The geodetic coordinates of the origin L_0, B_0, H_0 and its geodetic azimuth A_0 with reference to a certain direction are called the geodetic origin data, which are the coordinate datum for classical geodetic survey.

The ellipsoid orientation is closely related to the determination of the geodetic origin data. To position and orient the ellipsoid is to determine the geodetic origin data, and once the geodetic origin data are determined, the positioning and orientation of the ellipsoid will be completed. As shown in Fig. 7.5, L_0, B_0 of the geodetic origin P define the normal to the ellipsoid passing through this point, but the ellipsoid can still rotate and translate about the surface normal, which will not be completely fixed unless H_0 and A_0 are determined.

Mathematically, however, we determine the position and orientation; any set of L_0, B_0, H_0, and A_0 will enable us to define the relationship between the ellipsoid and the geoid, yet such a relationship not appropriate. The reference ellipsoid is a mathematical figure of the geoid and we should attempt to make it better approximate the regional geoid. This is the only way that the observed elements reduced to the ellipsoid can be practically significant. Meanwhile, the deflection of the vertical and the initial geodetic azimuth will also be conveniently solved. Hence, the positioning and orientation of the ellipsoid are required to satisfy the following conditions:

1. The minor axis of the ellipsoid is parallel to the Earth's rotation axis
2. The planes of the initial geodetic and astronomical meridians are parallel to each other
3. The ellipsoid surface best fits the geoid in an area of interest

The following analytical expressions can be used to illustrate these three conditions:

1. $\varepsilon_X = 0$, $\varepsilon_Y = 0$
2. $\varepsilon_Z = 0$ (based on $\varepsilon_X = 0$)
3. $\sum N^2 = \text{minimum}$

where $\varepsilon_X, \varepsilon_Y, \varepsilon_Z$ are the Euler angles and N is the geoid undulation.

With the above three conditions, if the first two conditions are satisfied (abbreviated as "double parallel"), the ellipsoid will be an approximation of the true shape of the Earth. The simplest formulae for the deflection of the vertical and the Laplace azimuth can be formed as:

$$\left.\begin{aligned} \xi &= \varphi - B \\ \eta &= (\lambda - L)\cos\varphi \end{aligned}\right\}, \tag{7.24}$$

$$A = \alpha - (\lambda - L)\sin\phi. \tag{7.25}$$

The third condition can ensure that the ellipsoid closely approximates the geoid, so the corrections applied to the reduction of observed quantities are fairly small and will better agree with the real quantities.

For the ellipsoid orientation, we can allow the defined L_0, B_0, H_0, and A_0 to satisfy these three conditions by employing the following methods.

Equations (7.24) and (7.25) are obtained under the condition of "double parallel," and L_0, B_0, A_0 defined by these two equations are:

$$\left.\begin{aligned} L_0 &= \lambda_0 - \eta_0 \sec\varphi_0 \\ B_0 &= \varphi_0 - \xi_0 \\ A_0 &= \alpha_0 - \eta_0 \tan\varphi_0 \\ H_0 &= H_{\text{Orthometric0}} + N_0 \end{aligned}\right\}. \tag{7.26}$$

If the defined geodetic origin data satisfy (7.26), it will meet the condition of "double parallel", i.e., the conditions 1 and 2 for determining the position and orientation. Equations (7.24) and (7.25) are the forms when (7.10) and (7.11) take $\varepsilon_X = \varepsilon_Y = \varepsilon_Z = 0$. Here, $\varepsilon_X, \varepsilon_Y, \varepsilon_Z$ determine the orientation of the ellipsoid, which are called the orientation parameters of the reference ellipsoid.

In (7.26), $\lambda_0, \varphi_0, \alpha_0$, and $H_{\text{orthometric0}}$ can be obtained by astronomical survey and leveling, and ξ_0, η_0, N_0 are deflections of the vertical and geoid undulation (geoid–ellipsoid separation) at the geodetic origin.

How to make the defined L_0, B_0, H_0, A_0 fulfill condition 3 of positioning and orientation is a matter of choice of the ξ_0, η_0, N_0. The role of ξ_0, η_0, N_0 is quite similar to that of $\Delta X_0, \Delta Y_0, \Delta Z_0$ in the Bursa–Wolf model (from (7.28), neglecting

$\varepsilon_X, \varepsilon_Y, \varepsilon_Z$, and so on, ξ_0, η_0, N_0 can be defined by $\Delta X_0, \Delta Y_0, \Delta Z_0$). These determine the position of the ellipsoidal center and are thereby called the positioning parameters of the reference ellipsoid.

Based on the different approaches for obtaining ξ_0, η_0, N_0, we can classify the orientation methods into single astronomical position datum orientation and astronomical–geodetic orientation. When a datum is oriented by a single astronomical point, we simply take $\xi_0 = 0, \eta_0 = 0, N_0 = 0$.

The above equations have shown that, at the geodetic origin, the direction of the normal to the ellipsoid coincides with that of the plumb line and the ellipsoid is tangent to the geoid. It follows from (7.26) that:

$$L_0 = \lambda_0, B_0 = \varphi_0, A_0 = \alpha_0, H_0 = H_{\text{orthometric0}}.$$

It can be seen that the single astronomical position datum orientation, in nature, is to consider the astronomical longitude, latitude, and azimuth measured at the geodetic origin as the geodetic longitude, latitude, and geodetic azimuth. The orthometric height (or normal height) of the geodetic origin is considered the geodetic height. Using this method it is difficult to make the ellipsoid fit the geoid within a large area (see, e.g., DMA 1984). Hence, after basically completing the national astro-geodetic survey, we tend to reorient by making use of the observational results on the condition that $\sum N^2 = \text{minimum}$. This is the astronomical–geodetic orientation.

The ellipsoid can be positioned and oriented by the method of astronomical–geodetic orientation, which provides arc measurement equations at multiple astro-geodetic points and obtains ξ_0, η_0, and N_0 by adjustment computations.

7.3.2 Arc Measurement Equation

Arc measurements can be categorized as ancient, neoteric, and modern.

In ancient societies, when people started realizing that the Earth is a sphere, they could then technically estimate the shape and size of the Earth by the length of an arc between two points and the measurement of difference in latitude at the two points. This was arc measurement in early times.

The first documented measurement of the size of the Earth was by the Hellenic scholar Eratosthenes (276–194 BC). He estimated that the radius of the Earth was 6,844 km. Since no field observations were carried out, this could not be considered an actual arc measurement. The first country to carry out an actual arc measurement was China. In 724 AD (Kaiyuan 12 years, Tang Dynasty), presided over by the Chinese astronomer Yixing (birth name Zhang Sui), the imperial astronomer Nan

Gongyue measured the distances between Huaxian, Junyi (now Kaifeng), Fugou, and Shangcai. He also measured the altitude of the North Pole in these four places and the shadow cast by the sun at midday on the summer solstice. He concluded that the length of 1° of meridian arc was 351 Li and 80 Bu (by Tang measurement, 1 Li equals 300 Bu). Since 1 Tang Li is equivalent to 1,500 Tang Chi, and 1 Tang Chi is equal to 24.75 cm, we find that 1°of arc length is 130.38 km. The ancient astronomers set a circumference of 365.25°, and it was converted to 360°, so 1°of arc length was calculated to be 132.28 km. This value, although 21 km larger than the given arc length of 1° being equal to 111 km, is extraordinary considering the level of technology back then (Xiong 1985).

Since Newton first claimed that the Earth's shape was an ellipsoid and Snell proposed the method of triangulation, the early eighteenth century ushered in a new era of arc measurements. The concept of arc measurements has been extended to determining the two elements of the Earth ellipsoid, i.e., the semimajor axis a and the flattening f. Since the early 1800s, surveyors from different countries have been engaged in a great number of arc measurements and have calculated many results for the Earth ellipsoid. It can be seen from the formula for meridian arc length in Chap. 5 that the meridian arc length is the function of a and e^2 (or f). Based on the surveying results of many segments of the meridian arc on the Earth, we can find the solution of a and f (or e^2) by means of the least squares method. The currently used arc measurement equation is derived from (7.23). In practice, the new ellipsoidal elements are obtained using astronomical, geodetic, gravimetric, and satellite surveying data based on the original old ellipsoid. As a result, the calculation of the new ellipsoid elements is actually a process of successive approximation. Let the elements of the old ellipsoid be a_{old} and f_{old}, and the elements of the new ellipsoid be $a_{new} = a_{old} + \mathrm{d}a$, $f_{new} = f_{old} + \mathrm{d}f$. The problem now becomes to find $\mathrm{d}a$ and $\mathrm{d}f$.

It can be written from the formula for the deflection of the vertical that:

$$\begin{aligned}\begin{bmatrix}\eta_{new}\\ \xi_{new}\\ N_{new}\end{bmatrix} &= \begin{bmatrix}(\lambda - L_{new})\cos B_{new}\\ \varphi - B_{new}\\ N_{new}\end{bmatrix}\\ &= \begin{bmatrix}(\lambda - L_{old})\cos B_{old}\\ \varphi - B_{old}\\ N_{old}\end{bmatrix} + \begin{bmatrix}-\mathrm{d}L\cos B_{old}\\ -\mathrm{d}B\\ \mathrm{d}N\end{bmatrix},\end{aligned} \tag{7.27}$$

where $\mathrm{d}N = \mathrm{d}H$. Substituting (7.23) into the above equation yields:

$$
\begin{bmatrix} \eta \\ \xi \\ N \end{bmatrix}_{new} = \begin{bmatrix} \dfrac{\sin L}{(N+H)} & -\dfrac{\cos L}{(N+H)} & 0 \\ \dfrac{\sin B\cos L}{M+H} & \dfrac{\sin B\sin L}{M+H} & -\dfrac{\cos B}{M+H} \\ \cos B\cos L & \cos B\sin L & \sin B \end{bmatrix}_{old} \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix}
$$

$$
+ \begin{bmatrix} -\sin B\cos L & -\sin B\sin L & \cos B \\ \sin L & -\cos L & 0 \\ -Ne^2\sin B\cos B\sin L & Ne^2\sin B\cos B\cos L & 0 \end{bmatrix}_{old} \begin{bmatrix} \varepsilon_X \\ \varepsilon_Y \\ \varepsilon_Z \end{bmatrix}
$$

$$
+ \begin{bmatrix} 0 \\ \dfrac{N}{M}e^2\sin B\cos B \\ N(1-e^2\sin^2 B) \end{bmatrix} \Delta m
$$

$$
+ \begin{bmatrix} 0 & 0 \\ -\dfrac{N}{(M+H)a}e^2\sin B\cos B & -\dfrac{M(2-e^2\sin^2 B)}{(M+H)(1-f)}\sin B\cos B \\ -\dfrac{N}{a}\left(1-e^2\sin^2 B\right) & \dfrac{M}{1-f}\left(1-e^2\sin^2 B\right)\sin^2 B \end{bmatrix}_{old} \begin{bmatrix} \mathrm{d}a \\ \mathrm{d}f \end{bmatrix}
$$

$$
+ \begin{bmatrix} (\lambda - L_{old})\cos B_{old} \\ \varphi - B_{old} \\ N_{old} \end{bmatrix}. \tag{7.28}
$$

The above is the generalized arc measurement equation, where the unknowns are $\Delta X_0, \Delta Y_0, \Delta Z_0$, $\varepsilon_X, \varepsilon_Y, \varepsilon_Z, \Delta m, \mathrm{d}a$, and $\mathrm{d}f$. In practical cases, the values of $\varepsilon_X, \varepsilon_Y, \varepsilon_Z$, and Δm are always left out according to conditions 1 and 2 of the ellipsoid orientation. The new ellipsoidal elements, positioning, and orientation values can be calculated using the above equation.

Such equations of arc measurement as (7.28) can be written at every astro-geodetic point of the astro-geodetic network. Solving the equations based on:

$$
\sum \left(\xi_{\mathrm{new}}{}^2 + \eta_{\mathrm{new}}{}^2\right) = \text{minimum}, \tag{7.29}
$$

or

$$\sum {N_{\text{new}}}^2 = \text{minimum}, \tag{7.30}$$

one can obtain the ellipsoidal elements that will be a best fit for an area of interest such as $a_{\text{new}} = a_{\text{old}} + \mathrm{d}a$, $f_{\text{new}} = f_{\text{old}} + \mathrm{d}f$, and the positioning elements of the new ellipsoid $\Delta X_0, \Delta Y_0, \Delta Z_0$. Substituting the solution back into (7.28) will produce the values of $\xi_{\text{new}}, \eta_{\text{new}}$, and N_{new} at any arbitrary astro-geodetic point, certainly including ξ_0, η_0, and N_0 at the geodetic origin.

Since $\xi, \eta,$ and N are correlated, theoretically (7.29) is equivalent to (7.30). However, it can be seen from (7.28) that, by changing the ellipsoidal elements, the value of η remains unchanged, which indicates that the variations in deflections of the vertical with the ellipsoidal elements are insignificant. Therefore, solving the arc measurement equation according to the conditions in (7.29) will yield a result of lower accuracy. Besides, considering that the change of N is milder than that of ξ, η, the effect of the local anomalies will be comparatively less. Thus, in practice, we usually allow $\sum N_{new}^2 = \text{minimum}$. When adopting the normal height system, correspondingly, the condition of $\sum {\zeta_{new}}^2 = \text{minimum}$ should be satisfied.

We have to point out that a nation, although with vast territory, is still confined to a small proportion of land with respect to the whole Earth. Hence, the ellipsoidal elements obtained by solving the arc measurement equations based on the surveying data from one country are often dramatically different from those based on data worldwide. One point is that the semimajor axis and flattening of the Earth ellipsoid based merely on China's astro-geodetic data were posited as approximately 6,378,670 m and 1:292.0, respectively. Therefore, in the establishment of the Xi'an Geodetic Coordinate System 1980 of China, these two parameters of the size of the Earth ellipsoid were left out. The a and f values adopted were those recommended at IUGG1975. In this case, solving the arc measurement equation becomes a matter of determining the position and orientation of the ellipsoid.

Hence, the astronomical-geodetic orientation means providing the arc measurement equations at the original astro-geodetic point, which can be written as:

$$\begin{aligned} N_{new} = {} & \cos B_{old} \cos L_{old} \Delta X_0 + \cos B_{old} \sin L_{old} \Delta Y_0 + \sin B_{old} \Delta Z_0 \\ & - \frac{N_{old}}{a_{old}} \left(1 - e_{old}^2 \sin^2 B_{old}\right) \Delta a + \frac{M_{old}}{1 - f_{old}} \left(1 - e_{old}^2 \sin^2 B_{old}\right) \sin^2 B_{old} \Delta f + N_{old}. \end{aligned} \tag{7.31}$$

Based on $\sum N_{new}^2 = \text{minimum}$, we calculate the difference in position between the old and new ellipsoid centers $\Delta X_0, \Delta Y_0, \Delta Z_0$, and substitute them into the following equation:

$$\begin{bmatrix}\eta\\ \xi\\ N\end{bmatrix}_{\text{new}} = \begin{bmatrix} \dfrac{\sin L}{(N+H)} & -\dfrac{\cos L}{(N+H)} & 0 \\ \dfrac{\sin B\cos L}{(M+H)} & \dfrac{\sin B\sin L}{(M+H)} & -\dfrac{\cos B}{(M+H)} \\ \cos B\cos L & \cos B\sin L & \sin B \end{bmatrix}_{\text{old}} \begin{bmatrix}\Delta X_0\\ \Delta Y_0\\ \Delta Z_0\end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 \\ -\dfrac{N}{(M+H)a}e^2\sin B\cos B & -\dfrac{M(2-e^2\sin^2 B)}{(M+H)(1-f)}\sin B\cos B \\ -\dfrac{N}{a}\left(1-e^2\sin^2 B\right) & \dfrac{M}{1-f}\left(1-e^2\sin^2 B\right)\sin^2 B \end{bmatrix}_{\text{old}} \begin{bmatrix}\Delta a\\ \Delta f\end{bmatrix}$$
$$+ \begin{bmatrix} (\lambda - L_{\text{old}})\cos B_{\text{old}} \\ \varphi - B_{\text{old}} \\ N_{\text{old}} \end{bmatrix},$$

This gives $\xi, \eta,$ and N at each astro-geodetic point, including at the geodetic origin, and eventually provides the new geodetic origin data.

The result of the astro-geodetic orientation indicates that, at the geodetic origin, the direction of the normal to the ellipsoid does not coincide with that of the plumb line, and the ellipsoid is no longer a tangent to the geoid. However, the ellipsoid surface is the best fit to the geoid over an area of interest.

When determining the ellipsoid orientation in an area (non-global) of interest based on $\sum N^2 = \text{minimum}$, the ellipsoid center will not coincide with the Earth's center of mass. Thus, we get a local orientation or non-geocentric orientation. The established coordinate system is called the local coordinate system or non-geocentric coordinate system.

Differing from the above neoteric methods for arc measurement, the concepts of arc measurement in modern times have greatly expanded. Integrating the gravity and spatial geodetic surveying data worldwide, the arc measurement in modern times studies the Earth from both geometric and physical perspectives, including the geometric shape and size of the Earth ellipsoid as well as the gravity field of the Earth.

Another four fundamental parameters are used to describe the Earth, a (semimajor axis of the ellipsoid), GM (the product of gravitational constant and the mass of the Earth), J_2 (second-order zonal harmonic coefficient of the Earth's gravity field), ω (angular velocity of the Earth's rotation), and a series of geometric and physical constants derived from these as well as the Earth gravity field model.

7.3.3 Significance of the Classical Method of Ellipsoid Orientation in Understanding the Principle of Establishing a Modern Geodetic Coordinate System

The principle of fixing the coordinate system to the Earth passing through surface points P and M in Fig. 7.5 is based on the precondition that the Earth is assumed to be rigid. The Earth is actually a non-rigid body, in fact it is a complex viscoelastic body. As a result, the surface points are subject to constant changes, which include both regular changes like the solid Earth tide and irregular changes like various unpredictable deformations. Hence, the coordinate systems cannot be accurately determined only from points P and M. Such indetermination can be improved by increasing the number of surface points used to define the coordinate system and by long-term repeated observations. The coordinates of these surface points should certainly not contradict each other; for example, the distance between points is subject to objective constraints. Then, we can use relative measuring techniques like GPS or VLBI to determine the distance between points.

It is thus clear that the chosen coordinate system fixed to the Earth is realized by determining the coordinates of a set of surface points (datum points). Alternatively, we might say that a self-consistent set of station coordinates contains (or determines) the coordinate system. These points are the VLBI, SLR, and GPS points. The fundamental principle of establishing a modern geodetic coordinate system is given above.

7.4 Conventional Terrestrial Reference System

7.4.1 The Geocentric Coordinate System and Its Application

A coordinate system with its origin at the center of mass of the Earth is called a geocentric coordinate system. Likewise, a coordinate system with its origin at the geometric center of the reference ellipsoid based on classical surveying techniques is called a non-geocentric coordinate system (also referred to as the local coordinate system). Identical with the local coordinate system, the geocentric coordinate system can also be categorized into geocentric geodetic coordinate system and geocentric geodetic Cartesian coordinate system. In modern geodesy, a geocentric coordinate system is usually referred to as a terrestrial reference system (TRS). The pointing direction of the coordinate axes of the TRS changes under the effect of the Earth's polar motion, which will present obstacles in practical use. Consequently, as early as 1967, the International Astronomical Union (IAU) and the International Association of Geodesy (IAG) recommended that the 1900–1905 mean latitude of the five International Latitude System stations be used as the datum point. The position of the mean pole corresponds to the mean position of the rotation axis over

the above period, which is called the Conventional International Origin (CIO). The corresponding Earth equatorial plane is referred to as the mean equatorial plane or conventional equatorial plane. In practice, CIO is commonly used as the Conventional Terrestrial Pole (CTP) up to the present time. The TRS that refers to the CTP is called the Conventional Terrestrial Reference System (CTRS). The TRS corresponding to the instantaneous pole at true-of-date is called the Instantaneous Terrestrial Reference System (see Moritz and Mueller 1987).

When being applied to the plotting of topographic maps and engineering on a regional basis, the geocentric coordinate system is not immediately required. However, in applications involving cross-regional surveying and mapping projects, integration as well as studies on space technologies, geodynamics, and the gravity field of the Earth that involve physical factors, the geocentric coordinate system plays a critically important role as described below.

Application Demands in Geodynamics, Physical Geodesy, and Space Technology

Following the 1960s, as an interdisciplinary subject, geodynamics has gradually drawn the attention of geodetic scholars. People have realized that almost all geodynamic phenomena, like crustal movement, material migration, tidal variation, and Earth rotation, can be monitored by applying space geodetic technologies. Therefore, geodesy has become one of the most fundamental methods for studying the geodynamic phenomena of the Earth. Broadly speaking, to carry out systematic studies on geodynamics, a fairly stable reference system should be established first. Otherwise, data observation will not be homogeneous. Hence, defining and realizing a reference system that is suitable for the purpose of geodynamic study becomes crucial.

Differing from the local coordinate system established in static geodesy, dynamic geodesy considers the Earth as a non-rigid, deformable body. Consequently, there arises the important issue of where to put the origin of the reference system within the Earth to maintain its stability. If the Earth is considered as a system of particles with the inside resultant force being zero, then, according to the dynamics of a system of particles, the Earth moves in a certain fixed orbit acted upon by the resultant external force of other celestial bodies. The law of motion of the Earth's particles completely corresponds to the hypothesis that all the mass of the Earth is concentrated in the center of mass. Therefore, however the mass inside the Earth migrates or how the shape of the Earth deforms, the Earth's center of mass moves along a fixed orbit. The significance of such an inference is that although the position of a point on the Earth's surface is changing at any moment, the position of the Earth's center of mass is "fixed" when observed from the outer space. Hence, with respect to the study of geodynamics, the ideal position of the origin of the reference system should be at the Earth's center of mass.

In physical geodesy, we need to choose a normal ellipsoid that best fits the geoid on a global basis, and its geometric center should coincide with the Earth's mass

center. Meanwhile, the gravity field needs to be represented using a reference system whose origin is at the center of the Earth (see Sect. 4.1).

The geocentric coordinate system plays a significant role in space technology. When artificial Earth satellites are orbiting the Earth, their orbital plane passes through the Earth's center of mass. As a result, the coordinate system of choice for orbital calculations is the geocentric system. Of course, tracking and observation of the satellites will also be in the coordinate system with the origin at the Earth's center of mass, otherwise their orbits will probably not be calculated accurately and tracking will be adversely influenced.

Application Demands in Surveying and Mapping Engineering

In surveying and mapping engineering, the demand of the customers for surveying and mapping products has undergone dramatic changes. People have realized that using geocentric datum can simplify application of the integrated geographic information (an interface switch is not needed) and can enable the geographic information to integrate "seamlessly" in terms of category, space, and time. The detailed interpretations are as follows:

The Demand for Integrated Application in Navigation and Construction Layout

Currently and in the foreseeable future, maps (including spatial database) are mostly utilized in association with the application of satellite navigation, such as vehicle-carried navigation systems, dynamic command and management, and so on. Using the geocentric datum as the mathematical basis for maps enables direct orientation of the space navigation results on the map platform using technologies like GPS. This differs from the traditional application demand whereby people are merely concerned with finding the relative position instead of the absolute position using maps; for instance, determining the position and orientation referenced to isolated features or residents. Naturally, users do not have a demand for geocentric datum. As well as the integration of various kinds of geographical information, the geographical information of different areas and at different times referenced to the same datum can be directly superimposed on each other, and thus an additional interface switch can be avoided.

Similarly, in engineering construction, GPS is employed in the course of actual measurements such as during construction layout. Hence, if the project design is referred to the geocentric coordinate system, the transformations of coordinate systems will be avoided.

The Demand for Holistic Geographical Information in the Spatial Domain

The Earth is the all and the one. Social and technological advancement will break the traditional pattern of applying geographical information only within national boundaries, and instead will spread it all over the geospace. Referring to the geocentric datum is fundamental for the realization of such a transformation.

The geocentric datum is the precondition for ensuring a "seamless" splice between nautical charts and topographic maps. The International Hydrographic Bureau (IHB) has adopted the WGS84 (this system has been consistent with the ITRF since 1994) as the datum for nautical charts. The practical significance of unifying the datum of nautical charts in all countries using a geocentric datum is evident in navigation. If the datum of topographic maps agrees with that of the nautical charts, it will be favorable for inshore ships and aerial navigation. Otherwise, additional external transformations will be needed between the navigation system and map platform as well as between the nautical charts and topographic maps.

The implementation of the transnational and intercontinental Geographic Information System (GIS) and global mapping programs launched by international organizations should refer to the geocentric system. High dynamic geodetic technologies can determine the instantaneous geocentric coordinates of the remote sensing platform, which indicates that the remote sensing information directly refer to the geocentric datum.

The Demand for Holistic Geographical Information in the Time Domain

The temporal database requires the reference datum to be maintained for a long time. The non-geocentric coordinate system is static whereas the geocentric datum can work long-term. In GIS, historical, current, and future information should refer to the same datum, which requires the datum system to be continuous in the time domain. The kinematic geocentric reference frame can simulate the position drift generated by the tectonic plates. Although involved in the application of cartography, this effect can be discounted within two to three decades, yet once it is taken into consideration the geocentric datum can ensure system continuity.

Furthermore, high-precision data of the control points in the digital age have all been stored in the spatial database. However, the precision of the control points on the manual analog maps are subject to the accuracy of plotting. This is to say that the digital maps will not negatively affect the use of high-precision datum.

The differential GPS technique has brought into reality the calculate-and-use facility of the control point. The geocentric system is referred to as the datum and is required to be continuous in time sequence, i.e., the points measured at different time periods should belong to the same system.

The Demand for Development of Geodetic Technology

GPS and some other analogous satellite positioning systems have pervaded or will pervade every aspect of human society. GPS employs the geocentric coordinate system, which requires geodetic measurement and its products to utilize its corresponding coordinate systems.

The dynamism of precise geodetic measurements requires that it be necessary to refer to the high-precision geocentric coordinate system with physical significance.

7.4.2 *Definitions of the CTRS and the Conventional Terrestrial Reference Frame*

Earth's Polar Origin and the Origin of Longitude

The Earth's pole is the intersection of the Earth's spin axis with the crust. The movement of the Earth's rotation axis with respect to its crust causes temporal variations in the position of the Earth's pole on the Earth's surface, which is called the polar motion of the Earth (shortened to polar motion). The time-varying spin axis of the Earth is the instantaneous axis, and the corresponding pole is called the instantaneous pole.

The movement of the instantaneous poles on the Earth's surface is contained within a small area, so we can use a plane tangent to the Earth's surface through the center of the pole to replace the Earth's surface within this area. Usually, a plane rectangular coordinate system will be established on this plane to determine the instantaneous position of the pole, so this coordinate system is called the polar coordinate system. We take the tangent point as the origin, denoted by P_0. This point is the mean position of the instantaneous pole during some time period (mean pole) and is defined as the Earth polar origin (cf. Fig. 7.6). Customarily, we take the direction of the Greenwich meridian passing through P_0 to represent the positive direction of the x-axis. The direction of the meridian 90° west of Greenwich represents the positive direction of the y-axis. Hence, the coordinates of the instantaneous pole P can be expressed by the rectangular coordinates (x_p, y_p).

Based on the polar motion matrix $A = R_Y(-x_p)R_X(-y_p)$ (cf. Fig. 7.2, y_p and x_p are the Euler angles ε_x, ε_y), we establish the relationship between the CTRS and the Instantaneous Terrestrial Reference System as:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{\text{CTRS}} = A \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{\text{ITRS}} \tag{7.32}$$

The organizations involved in determining polar motions have always included the International Latitude Service (ILS), the International Polar Motion Service

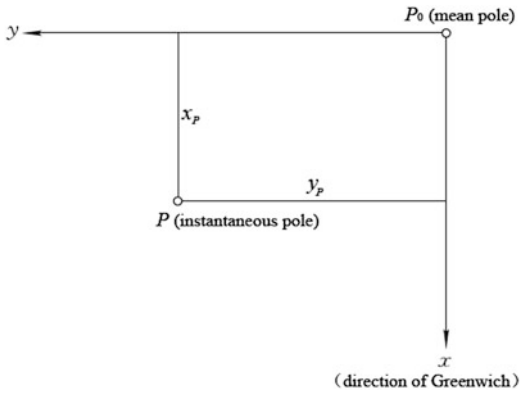


Fig. 7.6 Polar coordinate system

(IPMS), and the Bureau International De L'Heure (BIH). Since the 1970s, with the births of new technologies like Satellite Laser Ranging (SLR), Lunar Laser Ranging (LLR), Very Long Baseline Interferometry (VLBI), and the Global Positioning System (GPS), the traditional method of determining polar motion, optical astrometry, has been replaced by the state-of-the-art method of space geodesy. Hence, international organizations have decided to adopt the technologically advanced International Earth Rotation and Reference Systems Service (IERS) instead of organizations like IPMS and BIH.

For several decades, various polar systems have been created with gradually more in-depth studies on polar motion. At the 32nd symposium of the IAU in collaboration with IUGG, held in Stresa, Italy, 1968, the 1900–1905 mean latitude of the five International Latitude System stations was recommended to be the reference to define the position of the mean pole. This position of the mean pole relative to the 1900–1905 mean pole of the epoch (1903.0) is called the Conventional International Origin (CIO). ILS, IPMS, BIH, and other such organizations have successively used different optical instruments and mathematical processing approaches in an attempt to maintain this polar origin, so there are different CIO systems. The BIH systems to which CIO is subordinated include $BIH_{1968.0}$, $BIH_{1979.0}$, and $BIH_{1984.0}$, etc.

At present, the CIO system is maintained by the IERS. It provides users with data on the instantaneous pole through solution and regular publications based on the data obtained from the global observation stations; Fig. 7.7 illustrates the motion path of the instantaneous pole relative to the CIO from 1 July 2010 to 1 July 2013.

Unlike the geodetic latitude, where the equator is a natural origin, there is no natural starting point for the geodetic longitude. The standard reference meridian adopted to determine the geodetic longitude and universal time is called the initial meridian. The origin of longitude is defined by where the initial meridian and the equator intersect.

In 1884, astronomical delegates met at the international conference held in Washington to decide the meridian (line of longitude) that runs through the Observatory at Greenwich in southeast London, England as the initial meridian.

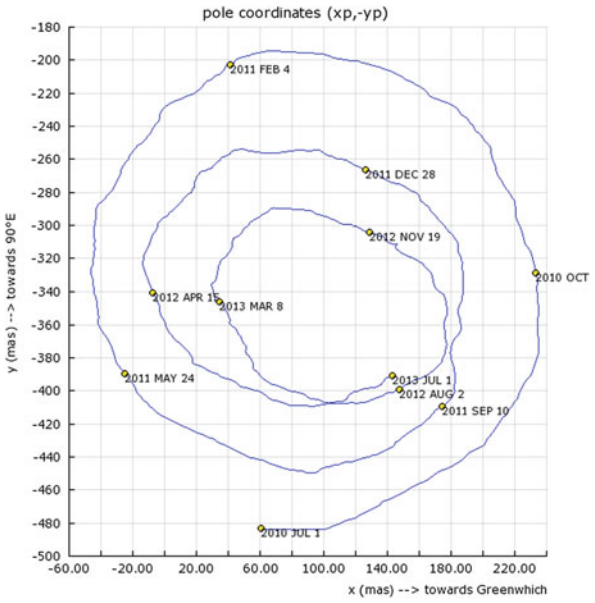


Fig. 7.7 The motion trajectory of the instantaneous pole relative to CIO from 1 July 2010 to 1 July 2013. Source: IERS

The point at which the prime meridian and the equator intersect is designated as the origin of longitude worldwide for longitude determination. Since 1957, a number of observatories with good long-term stability have been used internationally to maintain the longitude origin. Averaging the origin of longitude reversely obtained from the original observatory longitude values adopted gives the mean observatory longitude origin. After the CIO was defined as the Earth's polar origin in 1968, the meridian that passes through the CIO and the mean observatory longitude origin is called the initial meridian. The 0° longitude defined by various new technologies is made, as far as possible, coincident with the initial meridian, yet there still tends to be an error of less than 1″.

Reference System and Coordinate System

To describe the state of an object, such as determining the position of a particle, or describing the law of motion of a particle or body of mass (the Earth or a satellite), a reference system must be selected, without which discussions about laws of motion make no sense.

A coordinate system differs from a reference system in that it is always used to describe the mathematical representation and attaches more importance to the mathematical meaning, whereas a reference system is more concerned about the reference rather than the mathematical representation, and physical meaning is highlighted. For instance, establish an Earth-fixed reference system with its origin at the Earth's center of mass (terrestrial reference system). In this reference system

we can use Cartesian, spherical, or other coordinate systems to describe the laws of motion of a particle or a particle system. The origins of these coordinate systems are all at the geocenter and they are mutually static. Hence, the laws of motion of the particles described are completely identical and are independent of the choice of coordinates. Therefore, one reference system includes many coordinate systems that are essentially the same, such as geodetic Cartesian coordinates (X, Y, Z), geodetic coordinates (L, B, H), and Gauss plane coordinates (x, y), which are called equivalent or homogenous coordinate systems. Any arbitrary selection of these coordinates will exert no effect on the description of objective laws.

In geodesy, if a coordinate system is chosen, it suggests that the reference system is also selected. However, if a reference system is selected, any arbitrary coordinate system under this reference system is readily accessible. It is helpful to distinguish between the reference system and coordinate system at the conceptual level. When the established system is applied in surveying and mapping engineering to allow the geometric representation of the terrain, it is appropriate to call such a system the coordinate system. When applied in a context such as deformation monitoring or aerospace tests, with physical meaning involved, it will be more appropriate to call it a reference system. In practice, the reference system and coordinate system are often interchangeable with each other.

Ideal Terrestrial Reference System

The TRS is the reference system fixed to the Earth in some assured manner. If the Earth were an ideal rigid body, then any triaxial coordinate system fixed to the Earth could be well suited. The choice of the TRS merely depends on the user-friendliness. However, the Earth is really a deformable body, and there is a relative motion of parts of the crust with respect to each other. Hence, getting the coordinate system fixed to the Earth in an ideal way has become a matter of vital importance.

The bodily movement (such as polar motion) and the partial movement (like crustal deformation) of the Earth are superimposed upon each other in an observation. We can use an Earth-fixed reference system to separate them. The reference system should be defined by the following theoretical conception: the crust only deforms with respect to the reference system without rotation or translation as a whole, while the inertia reference system only entails bodily movement of the Earth such as revolution and self-rotation. This theoretical concept can be described by the following Tisserand condition:

$$\int_c \vec{v} dm = 0, \text{and} \int_c \vec{x} \times \vec{v} dm = 0,$$

where $\vec{x}$ and $\vec{v}$ are the position and velocity of a mass unit of the Earth dm with respect to the defined reference system. The domain of integration c is the entire Earth. The first part of the condition indicates the zero linear momentum of the Earth relative to the defined reference system and null translation of the Earth; the

second part of the condition shows that the angular momentum of the Earth with respect to the defined reference system is zero, indicating null rotation of the Earth. The reference system is defined in such a way that it separates the bodily and partial movement of the Earth, which is called the ideal TRS. If such a reference system can be realized, we will be justified in believing that the movement of the observed station described in this reference system belongs to the motion of the observed station itself.

In practical applications, the Tisserand condition is approximately realized by the no-net-rotation (NNR) condition.

Conventional Terrestrial Reference System

In order actually to establish the TRS, we have to choose some physical datum (like origin, scale, and orientation that involve physical meaning). The choice of the physical datum, however, is rather inadequate and arbitrary. A convention needs to be created in order to unify the datum. Hence, the terminology "convention" is used to describe this specific choice (see Pent and Luzum 2010), which is what the CTRS is all about.

The definition of the CTRS comprises:

1. Origin: the origin is at the geocenter, which is defined as the center of mass of the whole Earth, including oceans and atmosphere
2. Scale: the International System of Units (SI) meter is used for its unit of length, defined in a local Earth frame in the meaning of a relativistic theory of gravitation
3. Orientation: the orientation is consistent with that of the BIH system at 1984.0
4. Time evolution of the orientation: no residual global rotation is created with regards to the crust

Conventional Terrestrial Reference Frame

The reference frame is a realization of the reference system. Once the CTRS is chosen, it should be made accessible to all kinds of users. Hence, the system will be represented by a number of points on the Earth's surface. A globally-distributed set of 3D Cartesian coordinates infers the location of an origin, the orientation of an orthogonal set of Cartesian coordinate axes, and a scale parameter. In this sense, the Conventional Terrestrial Reference Frame (CTRF) is defined as a set of physical points with precisely determined coordinates in a specific coordinate system as a realization of the reference system. As a result, the well-known "geodetic control network" belongs to the concept of the reference frame.

The establishment of the terrestrial reference frame on the deformable Earth can be realized by transforming parameters and changing coordinates of the stations. The time-varying station coordinates will be determined repeatedly by the international service agency. All the global systematic errors can be reduced to scale, translation, or rotation errors and fall under the seven transformation parameters.

Therefore, under the constraint of the Tisser and condition, the transformation parameters calculated and published by the international service agency define the frame of the realized system.

7.4.3 Establishment and Maintenance of the CTRF

To go from definition to realization of a TRS, the following need to be attained:

1. Give the theoretical definition and define the conventions of the TRS (as described in Sect. 7.4.2).
2. Establish observation sites on the Earth's surface and carry out space geodetic measurements.
3. Process the observed data using the internationally recommended set of models and constants according to the conventions of the CTRS and solve the station coordinates of each observation site at a certain epoch; namely, establish the CTRF.
4. Analyze and deal with various kinds of deformation factors that influence the stability of the surface observation sites, and establish corresponding time-variant models to maintain stability of the CTRF.

To determine the position of the Earth's center of mass, let us first assume that the Earth is a rigid body. We can determine the distance between n points ($n \geq 3$) on the Earth's surface and the Earth's mass center using satellite dynamic techniques like SLR, followed by determination of the distances between these points by measuring techniques like GPS, VLBI, etc. Thus, we can determine the position of the geocenter through geometric constraint conditions. Nevertheless, the Earth is not a rigid body, it is a complicated viscoelastic body. Consequently, these points are changing all the time. The changes include both regular changes like the solid Earth tide and irregular changes like unpredictable deformations. The precise position of the Earth's center of mass is yet not determined and such indetermination can be improved through repeated observations over a long period. It is not difficult to understand that the accuracy of the position of the geocenter relative to the sites obtained using the method of least squares depends not only on the measuring precision, but also on the number of sites and the graphic structure.

The pointing direction of the coordinate axis of the TRS is related to the spin axis of the Earth. First, we have to clarify that the Earth's instantaneous axis of rotation is bound to pass through the Earth's center of mass. Otherwise, the center of mass will undoubtedly rotate around the instantaneous axis of rotation, which clashes with the kinetics of particles. Due to polar motion, the path of instantaneous axis of rotation forms an approximately circular conical surface, with its apex at the Earth's center of mass. Let the mean rotation axis (i.e., the axis of symmetry of the circular conical surface) be the Z-axis. The X-axis lies within the Greenwich Observatory's meridian plane. Then, choose the Y-axis, making O-XYZ a right-handed Cartesian coordinate system. In this case, we have established the geocentric reference

system. In practice the three axial directions of the reference system are determined by the Earth Rotation Parameters (ERP) provided by BIH or IERS.

The terrestrial reference frame (TRF) can be established and maintained by SLR techniques alone. Due to factors like variant models adopted, different numbers of sites, and different amounts of data used in the solution, the reference frames established by different SLR networks also differ to some degree.

The orientation and scale of the reference coordinate system can be determined precisely using the VLBI technique, yet this technique cannot determine the origin. Hence SLR is always used to determine a certain station coordinate as the initial point; for instance, the VLBI network uses the American Westford station as the initial point. Likewise, the reference systems established by different VLBI networks also differ from each other to some extent.

GPS and other technologies can also establish a TRF according to their own technological characteristics.

The International Terrestrial Reference Frame (ITRF) can be established by carrying out a combined adjustment between the above global SLR, VLBI, GPS networks, and other spatial geodetic networks.

The equation of the combined adjustment is:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{obs}} = \begin{bmatrix} \delta X \\ \delta Y \\ \delta Z \end{bmatrix} + \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{CTRF}} + \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix}, \tag{7.33}$$

where $[\delta X \quad \delta Y \quad \delta Z]^T$ is the deformation displacement of the observation station, $[X \quad Y \quad Z]^T_{\text{obs}}$ is the station coordinate vector corrected for translations (ΔX^0, $\Delta Y^0, \Delta Z^0$), rotations ($\varepsilon_X^0, \varepsilon_Y^0, \varepsilon_Z^0$), and scale correction (Δm^0) of the observed coordinates $\begin{bmatrix} X^0 & Y^0 & Z^0 \end{bmatrix}^T$ defined in the terrestrial reference frame with respect to the observational techniques "O" (such as SLR, VLBI, GPS, etc.), namely:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{obs}} = \begin{bmatrix} \Delta X^0 \\ \Delta Y^0 \\ \Delta Z^0 \end{bmatrix} + R_Z\left(\varepsilon_Z^0\right) \cdot R_Y\left(\varepsilon_Y^0\right) \cdot R_X\left(\varepsilon_X^0\right) \begin{bmatrix} X^0 \\ Y^0 \\ Z^0 \end{bmatrix} + \Delta m^0 \begin{bmatrix} X^0 \\ Y^0 \\ Z^0 \end{bmatrix}. \tag{7.34}$$

Equations (7.33) and (7.34) are the observation equations used to realize CTRF. The unknown parameters in the equations are $[X \quad Y \quad Z]^T_{\text{CTRF}}$ and $[\delta X \quad \delta Y \quad \delta Z]^T$, which define the CTRF. $\Delta X^0, \Delta Y^0, \Delta Z^0$, $\varepsilon_X^0, \varepsilon_Y^0, \varepsilon_Z^0$, and Δm^0 represent the relationship between CTRF and the terrestrial reference frame relative to the observation techniques "0".

ITRF is an example of CTRF. In practical applications there will usually be specific requirements of origin, scale, and orientation for the newly established ITRF.

According to the definition of CTRS, the coordinate origin of ITRF is situated at the center of mass of the whole Earth. As a dynamic technique, SLR can determine

the Earth's center of mass with fairly high precision. Thus, the coordinate origin of ITRF can be determined by this technique, i.e., it can let the translation parameter of a certain SLR network or the weighted sum of the translation parameters of several SLR networks involved in the adjustment be zero.

The scale of ITRF, by definition, is that of the local Earth frame within the meaning of a relativistic theory of gravitation. In practice, the scale is defined by physical parameters like the speed of light c, the gravitational constant of the Earth GM, and a certain model of relativistic correction adopted in data processing by the analysis centers. If the above values and models are different, the scale of each network will also be different. Currently, most analysis centers have adopted the values recommended by IERS, so the difference in scale between various networks is insignificant. The new ITRF scale is usually the scale of a certain network or the weighted mean of the scales of several networks.

As to the orientation of ITRF, the orientation parameters are usually considered to have known values.

The realization of ITRF of the present time, such as ITRF2000, ITRF2005, and ITRF2008, in terms of their input data, mathematical models, solution parameters, and solution methods is far more complex than that of earlier ITRFs. The underlying concepts, however, are very much alike. To understand better the methods of realizing the origin, scale, and orientation, we present here the establishment of the early ITRF91 (which applied GPS data for the first time) as an example to illustrate the process of ITRF realization.

Table 7.1 lists all the 16 networks involved in the establishment of ITRF91. The serial numbers 1–5 are the VLBI networks, 6–12 are SLR networks, 13–15 are LLR networks, and 16 is a GPS network. ITRF91 was eventually established by carrying out combined adjustment of all the data (see Boucher et al. 1992).

In combined adjustment, the reference system is defined as: the origin of the reference system as determined by SLR, the scale as determined by SLR coincident with VLBI, and the orientation, which is consistent with ITRF90.

The methods for realizing this reference system by combined adjustment are:

1. The transformation parameters of SSC (CSR) 92L01 network (Table 7.1, serial No. 6) are all assigned to be zero, i.e., an adjustment in the reference system is carried out to make the origin, scale, and orientation consistent with SSC (CSR). This result is defined as the ITRF91A. Apparently, the origin and scale of ITRF91A have satisfied the definition of ITRF91, but not the orientation.
2. The seven transformation parameters of ITRF91A and ITRF90 are solved to obtain the orientation correction parameters (cf. Table 7.2).
3. Both the translation parameters and scale factors of SSC (CSR) are set to zero and the rotation angles are R1, R2, and R3 as shown in Table 7.2 (cf. Table 7.1 serial No. 6). The transformation parameters obtained after the combined adjustment of all the data can be seen in Table 7.2.

For the datum definition adopted by ITRF*yy* (where *yy* is the annual series of this frame, denoting the last year whose data were used in the realization of the frame) at different dates, see Table 7.3 (Boucher et al. 1999, 2004; Altamimi et al. 2001,

Table 7.1 The networks in the combined adjustment of ITRF91 and their transformation parameters (translation parameters ΔX^0, ΔY^0, ΔZ^0, Euler angles $\varepsilon_X^0, \varepsilon_Y^0, \varepsilon_Z^0$, and scale factor Δm^0)

No.	Network	ΔX^0 (cm)	ΔY^0 (cm)	ΔZ^0 (cm)	Δm^0 (10^{-8})	${\varepsilon_X}^0$ (0.001″)	${\varepsilon_Y}^0$ (0.001″)	${\varepsilon_Z}^0$ (0.001″)
1	SSC(GSFC)92R03	0.3	−1.2	−2.8	−0.05	1.1	1.2	−2.5
2	SSC(NOAA)92R01	2.1	0.4	−2.2	0.08	4.6	9.9	−0.2
3	SSC(USNO)92R01	1.9	−4.4	0.8	−0.42	−0.4	−0.2	−0.2
4	SSC(NAOMZ)92R01	3.3	−2.9	−4.8	0.06	−1.3	−1.0	−0.5
5	SSC(JPL)92R01	−1.9	1.1	−1.7	−0.50	2.5	1.1	0.7
6	SSC(CSR)92L01	0.0	0.0	0.0	0.00	−0.4	0.4	−0.9
7	SSC(GSFC)92L01	−2.9	−1.4	0.3	0.31	−0.3	5.7	−7.3
8	SSC(DGFII)92L01	0.7	−3.0	0.0	−0.37	−284.4	−44.5	5.7
9	SSC(DUT)92L01	0.1	0.6	−3.8	−0.65	0.8	1.7	−4.6
10	SSC(GFZ)92L01	−0.1	−0.2	3.0	0.06	0.1	0.1	−0.2
11	SSC(GAOUA)92L01	−1.4	0.4	−7.2	−0.52	0.3	0.4	−1.4
12	SSC(NAL)92L01	0.4	−0.6	−3.5	0.08	1.0	1.6	0.6
13	SSC(UTXMO)92M01	−0.1	−3.5	17.6	−3.20	−6.5	0.8	38.5
14	SSC(JPL)92M01	−6.5	−1.2	4.7	−1.97	−1.0	1.4	−44.7
15	SSC(SHA)92M01	−6.8	−0.5	1.9	−2.75	−0.3	0.1	3.2
16	SSC(JPL)01P01	−6.7	14.8	−17.0	−0.14	−0.4	−0.7	61.5

Table 7.2 Transformation parameters (translation parameters $\Delta X_0, \Delta Y_0, \Delta Z_0$, Euler angles $\varepsilon_x, \varepsilon_y, \varepsilon_z$, and scale factor Δm) between ITRF91A and ITRF90

ΔX (cm)	ΔY (cm)	ΔZ (cm)	Δm (10^{-8})	ε_X (0.001″)	ε_Y (0.001″)	ε_Z (0.001″)
−0.1	0.4	1.5	−0.02	−0.4	0.4	−0.9

2002a, b, 2007, 2011). It can be seen from the table that the origin and scale of ITRF have been defined by the average of the chosen SLR, GPS, and VLBI solutions since BTS87 (where BTS is the reference system established by BIH). The orientation of ITRF was aligned to the BIH EOP (Earth Orientation Parameter) series, while the orientation of ITRF93 was aligned to the IERS EOP series. For the orientation time evolution, no global velocity field was estimated prior to ITRF90, so the AMO-2 plate motion model was recommended. Starting with ITRF91, the orientation time evolution satisfies the NNR condition with respect to (represented by "wrt" in Table 7.3) the previous ITRF series or the NNR-NUVEL1, NNR-NUVEL1A plate motion models. However, the orientation time evolution of ITRF93 was aligned to the IERS EOP series.

CTRF is four-dimensional or dynamic. Its dynamic feature lies in that CTRF is composed of coordinates and velocities at a reference epoch. Temporal variation (dynamism) is the innate property of CTRF. On the one hand, the plate motion, crustal deformation, and geodynamic factors as such will cause the instantaneous position of the point on the surface of the solid Earth to change constantly, and the station coordinates and station velocities should be provided. On the other hand, the new observational data are continually updated and new observation sites are ever

Table 7.3 Datum definition adopted by ITRF*yy*

ITRF solutions	Origin	Scale	Orientation	Time evolution of the orientation
ITRF0	BTS87	BTS87	BIH EOP	–
ITRF88	ITRF0	ITRF0	ITRF0	Recommended AMO-2
ITRF89	SLR[SSC(CSR)]	SLR[SSC(CSR)]	ITRF88	Recommended AMO-2
ITRF90	SLR[SSC(CSR)]	SLR[SSC(CSR)]	ITRF89	Recommended AMO-2
ITRF91	SLR[SSC(CSR)]	SLR[SSC(CSR)]	ITRF90	NNR wrt NNR-NUVEL1
ITRF92	SLR[SSC(CSR)]	SLR[SSC(CSR)]	ITRF91	NNR wrt NNR-NUVEL1A
ITRF93	SLR	SLR	IERS EOP	IERS EOP series
ITRF94	SLR + GPS	SLR + GPS + VLBI	ITRF92	NNR wrt NNR-NUVEL1A
ITRF96	SLR + GPS	SLR + GPS + VLBI	ITRF94	NNR wrt ITRF94
ITRF97	SLR + GPS	SLR + GPS + VLBI	ITRF96	NNR wrt ITRF96
ITRF2000	SLR (origin rate definition included)	SLR + VLBI (scale rate definition included)	ITRF97	NNR wrt NNR-NUVEL1A
ITRF2005	SLR	VLBI	ITRF2000	NNR wrt ITRF2000
ITRF2008	SLR	VLBI + SLR	ITRF2005	NNR wrt ITRF2005

increasing. Hence, new computational results of coordinates and velocities are regularly published.

The maintenance of CTRF, namely the retention of its dynamic features, primarily depends on the velocity field in actual measurement; for instance, the GPS Continuously Operating Reference System (CORS) network has a large amount of continuous observational data, which can be used to compute the station velocities and to provide observational data for maintenance of the TRF.

7.4.4 International Terrestrial Reference Frame and The World Geodetic System 1984

ITRF

The Establishment of ITRF

The International Terrestrial Reference Frame (ITRF) is a realization of the International Terrestrial Reference System (ITRS). The definition of ITRS is coincident with that of CTRS. The ITRF was established through a set of station coordinates (SSC) and station velocities derived from observations using space geodetic

techniques such as VLBI, SLR, LLR, GPS (since 1991), and DORIS (since 1994; Doppler orbitography and radio-positioning integrated by satellite). IERS regularly provides annual IERF solutions, which are published in the IERS annual reports. The available ITRF solutions include ITRF0, ITRF88, ITRF89, ITRF90, ITRF91, ITRF92, ITRF93, ITRF94, ITRF96, ITRF97, ITRF2000, ITRF2005, and ITRF2008. The basic steps to compute ITRF are as follows:

First, each individual SSC solution provided by the analysis centers is reduced to a reference epoch t_0 using their respective station velocity models; then, the method of combined adjustment yields ITRF station coordinates and seven transformation parameters for each SSC with respect to combined solutions of ITRF. The velocities of ITRF stations can be calculated by two methods: one is similar to calculating station coordinates except that its model is deduced from a coordinate transformation formula; the other is obtained by differentiating coordinates at two different epochs.

To represent the ITRF station coordinates by geodetic coordinates, IERS recommended the fundamental geodetic constants of the universal GRS (Geodetic Reference System). The currently adopted GRS80 was recommended by IUGG1979 and its ellipsoid parameters (see Table 4.1).

ITRF and IGS

With the establishment of the International Global Navigation Satellite System Service (IGS), ITRF and GPS are more closely interrelated (Altamimi and Collilieux 2009). IGS is in close collaboration with ITRF. On the one hand, ITRF provides IGS with an absolute long-term datum; IGS, on the other hand, provides global GPS observation data and helps to improve ITRF solutions.

Assuming that the reference epoch of a GPS campaign is T_c, we are allowed to make use of the observational data and precise ephemeris of this campaign to calculate the station coordinates. We might encounter two kinds of precise ephemeris; one is based on the WGS84 coordinate system while the other is based on ITRF (such as IGS precise ephemeris). The station coordinates obtained by using the precise ephemeris with different coordinates will end up in different coordinate systems.

When using the IGS precise ephemeris (whose reference frame is assumed to be ITRF*yy*), we should adopt the coordinates of given points in the ITRF*yy* reference frame (at reference epoch T_0), and reduce the coordinates from T_0 to the epoch of observation T_c based on the velocities of the given points in ITRF*yy*. If necessary, we can impose constraints on the reduced coordinates of the given points and, in this case, the coordinates of the unknown points obtained belong to the ITRF reference frame adopted by the IGS precise ephemeris.

Roles of ITRF in the Establishment and Maintenance of a Regional Geodetic Coordinate System

A geocentric coordinate system, by definition, is unique. However, different methods and data used in the realization will result in different geocentric coordinate systems. Since the geocenter of ITRF is highly precise, globally distributed, and considerably authoritative, other geocentric systems are also aligned to it. For instance, the WGS84 has been refined many times and the European reference frame has been incorporated into ITRF. ITRF station coordinates are strongly constrained by countries worldwide while processing their own GPS data, so as to make the national coordinate system approximate or pertain to the ITRF.

At present, the accuracy of ITRF station positions is at a level better than 1 cm and that of the station velocities is better than 3 mm per annum. We take these points as the initial stations and use GPS relative positioning to carry out GPS observations in the area of interest. After data processing, we will obtain the high-precision station coordinates in this area, which means that the geocentric coordinate system based on the technique of GPS has been established in this region of interest.

ITRF is crucially important in the establishment and maintenance of regional coordinate systems.

First, the precise IGS ephemeris and Earth orientation parameters (EOP) are used in establishing regional coordinate systems, and the reference frame of the IGS precise ephemeris belongs to the ITRF.

Second, the initial stations used in the establishment of regional coordinate systems are ITRF stations. In calculation, these stations are mostly tightly constrained. In this sense, the established coordinate system will be made well aligned to the ITRF. Therefore, the regional coordinate systems established in recent years have all made clear that their coordinate systems agree with ITRF. For instance, the European reference frame is fixed to a stable part of the Eurasian Plate and is coincident with ITRF at the reference epoch of 1989.0.

Although ITRF stations are chosen as the initial stations in the establishment of all regional geocentric coordinate systems, the approaches for selection are dramatically different, either by strongly constraining ITRF stations within and surrounding the area of interest (like the South American reference frame SIRGAS), or by constraining ITRF stations fixed to the stable part of different plates (such as the European reference frame). If different initial stations are chosen, the reference systems established will be different and such differences are often systematic.

WGS84

Since the 1960s, in order to establish an internationally unified geodetic coordinate system, the former American Defense Mapping Agency (DMA) established the WGS60, followed by the updated WGS66 and WGS72. In the mid-1980s the WGS84 coordinate system was established, which was a CTRS. Meanwhile,

WGS84 also includes the reference ellipsoid, fundamental constants, an Earth gravity field model, and a global geoid model (see NIMA 2000; NGA 2002).

The WGS84 reference frame was realized by a globally distributed set of monitoring stations. To establish the WGS84 reference system that agreed with ITRF, the USA refined WGS84 three times in 1994, 1996, and 2001, respectively. A refinement to the WGS84 in 2001 included data from 49 IGS stations and the station coordinate set has been given the designation WGS84 (G1150) at the epoch 2001.0. It was aligned to ITRF2000 with an accuracy of 1 cm in any coordinate component. Hence, for most applications that require an accuracy of less than centimeter level, WGS84 can be considered the same reference frame as ITRF.

For the four fundamental constants of the WGS84 Ellipsoid see Table 4.1.

7.5 Geodetic Coordinate Systems in China

7.5.1 Beijing Coordinate System 1954

The Beijing Geodetic Coordinate System 1954 (abbreviated as BJS54) is virtually the extension of the former Soviet Union Pulkovo Coordinate System 1942 in China. Thus, the latter will be discussed first.

Prior to 1946, the former Soviet Union adopted the Pulkovo Coordinate System 1932. Its geodetic origin was located in the center of the Great Round Hall at the Pulkovo Astronomical Observatory. It used the Bessel ellipsoidal parameters, oriented by a single astronomical position, only the initial geodetic azimuth was obtained through adjustment by using 45 Laplace's azimuths in the astro-geodetic network.

Such a coordinate system was not suitable for the former Soviet Union. First, the semimajor axis of the Bessel ellipsoid was too small, with an error of approximately 800 m. Second, the positioning and orientation of the reference ellipsoid was not well suited. In the Far East regions, the deviation of the geoid from the reference ellipsoid was up to 410 m. Therefore, the Pulkovo Coordinate System 1942 was established based on the Pulkovo Coordinate System 1932, but using parameters of the Krassowski Ellipsoid. The geodetic origin was still at Pulkovo, but the astro-geodetic orientation was performed this time.

During the 1950s, in the early phase of the establishment of an astro-geodetic network in China, to accelerate the cause of surveying and mapping and fully develop mapping, we were in desperate need of a geodetic coordinate system. Thus in 1954, with the advice and support provided by the parties concerned, and considering the historical condition at that time, the Surveying and Mapping Bureau of the General Staff Headquarters of the Chinese People's Liberation Army (SMBGSH) connected the first-order triangulation chains in China with the first-order chains in the Far East of the former Soviet Union. Using the former Soviet Union 1942 Pulkovo coordinates at the shared part of two baseline networks in

Huma and Jilalin and two triangulation points in Suifen River District as the initial data, and carrying out an adjustment of the first-order triangulation chains in northeast and east China, the coordinate system achieved by this transfer computation was designated as the Beijing Coordinate System 1954. So, it follows that the BJS54 is the extension in China of the Pulkovo Coordinate System 1942 of the former Soviet Union. However, this extension is slightly different from the Pulkovo Coordinate System 1942 in the strict sense; for instance, the height anomaly was estimated by astro-geodetic leveling in China based on the re-adjusted geoid in former Soviet Union 1955 as the initial value. The geodetic height is referenced to the Huanghai mean sea level obtained at the Qingdao tidal station, China, in 1956.

Highlights of BJS54 are summarized as follows:

1. It is a local coordinate system.
2. It has adopted the Krassowski Ellipsoid parameters (see Table 4.1).
3. It employs astro-geodetic orientation. η_0 and ξ_0 are obtained from 900 points (in the former Soviet Union) according to $\sum_{1}^{900}\left[\left(\eta-\eta_g\right)^2+\left(\xi-\xi_g\right)^2\right]=\text{minimum}$, where η_g and ξ_g are the components of the gravimetric deflection of the vertical obtained by the gravity method. The geoid separation (geoid undulation) of the geodetic origin N_0 is obtained from 43 points (selected evenly in the former Soviet Union astro-geodetic network) based on $\sum_{1}^{43}N^2=\text{minimum}$.
4. No Euler angles exist: $\varepsilon_X=\varepsilon_Y=\varepsilon_Z=0$.
5. The geodetic origin is located at Pulkovo in the former Soviet Union.
6. The height anomaly is estimated by astro-geodetic leveling in China, based on the re-adjusted geoid in former Soviet Union 1955 as the initial value.
7. The results of geodetic points are achieved by local adjustments after the establishment of BJS54.

The Beijing coordinate system 1954 had a tremendous impact on surveying and mapping. One hundred and fifty thousand national geodetic points and thousands of military control points, artillery survey control points, and mapping control points are all calculated based on this coordinate system. The surveying and mapping results, documents, and data based on BJS54 have been applied in many fields of economic construction and national defense building. In particular, the mission of preparing 1:50,000 scale and 1:100,000 scale topographic maps plotted using BJS54 is almost accomplished. The 1:10,000 scale topographic maps are also nearly completed.

The deficiencies and problems of BJS54 are as follows:

1. The Krassowski Ellipsoid parameters are different from the modern accurately defined ellipsoidal parameters, with its semimajor axis being approximately 108 m longer (compared to GRS80).
2. It involves only two ellipsoidal parameters with geometric properties (a, f) and cannot fulfill the needs of the four fundamental parameters (semimajor axis a, Earth gravity field second-order zonal harmonic coefficient J_2, geocentric

gravitational constant GM, and angular velocity of the Earth rotation ω) used to describe the Earth ellipsoid in both theoretical studies and practical applications.

3. The Krassowski Ellipsoid is adopted in geodetic surveying and computation, whereas in gravity data processing, the formula for normal gravity derived by Helmert in 1901–1909 is adopted:

$$\gamma_0 = 978030\left(1 + 0.005302\sin^2\varphi - 0.000007\sin^2 2\varphi\right).$$

   Hence, we cannot incorporate the two geometric parameters a and f of the Krassowski Ellipsoid with the physical parameters $\gamma_e = 978030$ mGal, and $\beta = 0.005302$ of Helmert's formula for normal gravity, used as the unified parameters for both geometric and physical geodesy.
4. The reference ellipsoid surface corresponding to BJS54 has a systematic slope from west to east relative to the geoid in China. In eastern regions the height anomaly can reach +65 m, and the nationwide average is 29 m (cf. Fig. 7.8).
5. It has ill-defined orientation. The pointing direction of the minor axis of the ellipsoid is neither the universally adopted CIO nor the polar origin $\mathrm{JYD}_{1968.0}$ in China. The plane of the initial geodetic meridian is not parallel to the Greenwich Mean Astronomical Meridian defined by the BIH. This is inconvenient and can cause errors in coordinate transformations.
6. Compared to the Xi'an Coordinate System 1980 in China, BJS54 has carried out a local rather than an integrated adjustment. The former, on the other hand, uses integrated adjustments and thus boasts higher accuracy than BJS54. The results of geodetic points provided stepwise by local adjustments will inevitably lead to some contradictions or unjustified situations.
7. The coordinate system is not established by China alone, and the geodetic origin is not in Beijing. So the "Beijing coordinate system" is more superficial than real, which may easily cause misperceptions. Its definition is neither simple nor clear enough.

   We should certainly note that these problems arose for historical reasons which, for a country just starting to establish an astro-geodetic network, was very hard to avoid.

7.5.2 *China's National Geodetic Coordinate System 1980 (Xi'an Coordinate System 1980)*

Long before the early 1960s, when the national astro-geodetic survey began to take shape, experts and scholars had started to embark on research by making use of the astro-geodetic survey data to calculate the ellipsoid parameters that better approximate the geoid in China. This attempt to tackle the deficiencies and problems of BJS54 reaped some early results.

In April 1978, the "Symposium of National Astro-Geodetic Network Integrated Adjustment" was held in Xi'an, China. Experts and scholars devoted a full

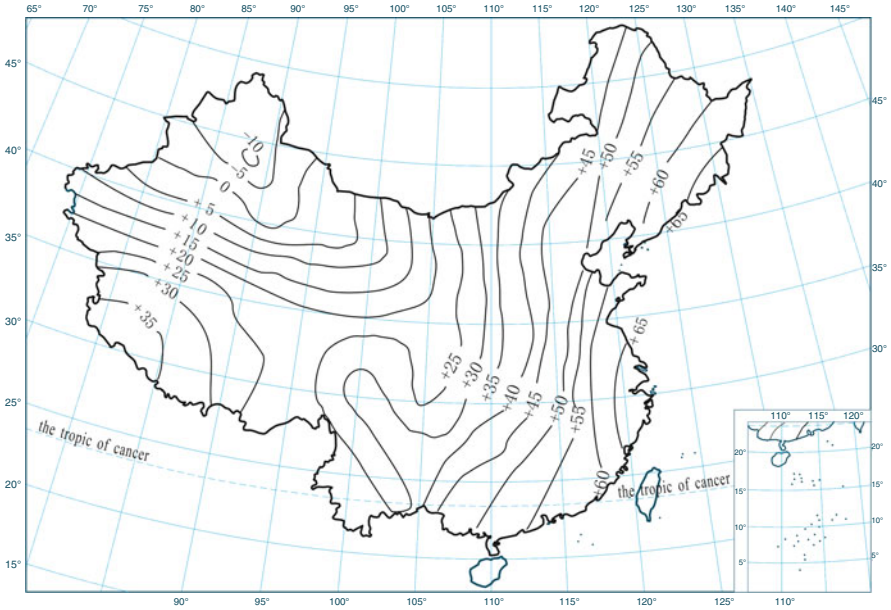


Fig. 7.8 China's mainland geoid, Beijing Coordinate System 1954

discussion and study to the issue of establishing a new geodetic coordinate system in China, holding that BJS54 had some technical shortcomings. Establishing China's own brand-new geodetic coordinate system was imperative and timely. In the conference summary, the following principles were defined concerning the issue of establishing China's geodetic coordinate system:

1. The integrated adjustments of the national astro-geodetic network will be carried out on a new surface of a reference ellipsoid. Thus, a new geodetic coordinate system should first be established, which was named the National Geodetic Coordinate System 1980 (later in practical use it is always referred to as the Xi'an Coordinate System 1980, abbreviated as XAS80).
2. The geodetic origin of the coordinate system is located at the center of China, close to Xi'an. The exact place is in Yongle township, Jingyang county, Shaanxi province.
3. The Earth ellipsoid parameters recommended by the IUGG in 1975 integrated the latest data throughout the world. The XAS80 has adopted four basic parameters (a, J_2, GM, and ω), and has calculated the Earth's flattening, normal gravity value at the equator, and coefficients of the normal gravity formula based on the four parameters.
4. The minor axis of the ellipsoid XAS80 is defined parallel to the direction from the Earth's center of mass to the polar origin $\mathrm{JYD}_{1968.0}$, and the initial geodetic

meridian plane should be parallel to the Greenwich Mean Astronomical Meridian.

5. The reference ellipsoid is oriented to minimize the sum of the squared height anomalies within China.
6. Taking into consideration the different needs of classical geodetic surveys and space geodetic surveys, and basing our decisions on the principle of "independence and self-setting, secure and easy access", we have established two sets of coordinate systems, respectively; namely, the XAS80 and the geocentric coordinate system. The former belongs to the local coordinate system according to the conditions for determining orientation. We have to maintain its long-term stability for the use of various departments throughout the country. The latter, based on the XAS80, has transformed the coordinate transformation parameters accurately obtained to the geocentric coordinates to satisfy the needs of developing long-distance weapons and space technologies. The geocentric coordinate transformation parameters can be refined with the incessant progress of surveying and mapping technologies and by comprehensive utilization of the ever-increasing astronomical, geodetic, gravimetric, and satellite geodetic survey data.

After the conference, the departments concerned established China's National Geodetic Coordinate System 1980 in line with the above principles.

XAS80 was established based on the BJS54. In terms of principles for its establishment, see Sect. 7.3.

The arc measurement equation used is (7.31). Due to the adoption of the normal height system, we therefore replace N with ζ, and replace Δf with Δe^2, which gives:

$$\begin{aligned}\zeta_{1980} = {} & \cos B_{1954}\cos L_{1954}\Delta X_0 + \cos B_{1954}\sin L_{1954}\Delta Y_0 + \sin B_{1954}\Delta Z_0 \\ & -W_{1954}\Delta a + \frac{1}{2}N_{1954}\sin^2 B_{1954}\Delta e^2 + \zeta_{1954}.\end{aligned} \tag{7.35}$$

We have adopted the ellipsoidal parameters recommended by IUGG1975, and thus $\Delta a = a_{\text{IUGG1975}} - a_{\text{Krassovsky}}$ and $\Delta e^2 = e^2{}_{\text{IUGG1975}} - e^2{}_{\text{Krassovsky}}$ are both known values.

The method for solving ζ_{1954} is as follows: We form 21 loops throughout the country using the results of 1,167 astronomical points and approximately 150,000 gravity points along the astro-gravimetric leveling lines, short-side astronomical leveling lines, and astronomical leveling lines, with isostatic corrections applied. We carry out an unequal-weighted adjustment to obtain differences in height anomalies along different leveling lines. We take ζ at the origin as the initial value to calculate one by one, and produce the national height anomaly map ζ_{1954}.

We select 922 points evenly nationwide at a $1^\circ \times 1^\circ$ interval and form an arc measurement equation like (7.35). According to $\Sigma_1^{922}\zeta_{1980}^2 = \text{minimum}$, solve the orientation elements ΔX_0, ΔY_0, and ΔZ_0, and then substitute them into (7.28) to produce the values of $\eta_{0(1980)}$, $\xi_{0(1980)}$, and $\zeta_{0(1980)}$ at the geodetic origin. According to (7.26), based on the astronomical longitude λ_0, latitude φ_0, and azimuth α_0 from

the origin point to another point and the normal height H_{normal0} at the geodetic origin, we can obtain $L_{0(1980)}$, $B_{0(1980)}$, $A_{0(1980)}$, and $H_{0(1980)}$ as the geodetic origin data of the XAS80.

XAS80 has adopted $\text{JYD}_{1968.0}$ instead of CIO as the origin of the polar coordinate (where JYD is the acronym of Chinese Pinyin *Ji Yuan Dian*, meaning polar origin). This polar coordinate system was established in 1977 by the Polar Motion Collaboration Group in China through analysis and study of the long-term and periodic components of the Earth's pole using the latitude survey data from the optical instruments of 36 sites abroad over the years 1949–1977. Participants of the Collaboration Group include the Department of Astronomy, Nanjing University, Shanghai Astronomical Observatory, Shaanxi Astronomical Observatory, Wuchang Time Observatory, and Tianjin Latitude Observation. The system has been maintained by Tianjin Latitude Observation using follow-up latitude survey data from basically the same sites and instruments, and the same methods for mathematical data processing.

We have to acknowledge that the precision of $\text{JYD}_{1968.0}$ is fairly high in terms of determining a polar system by means of optical techniques. The system is also quite stable. Its internal accuracy is equivalent to the CIO of BIH, but the average deviation of the external consistency is less than 0. 02″ compared with BIH and IPMS.

Although $\text{JYD}_{1968.0}$ is used in XAS80 in China, there have always been disputes over its employment in Chinese activities in the fields of astronomy and geodesy. In particular, with the improvements in space geodetic techniques, the traditional optical astrometric methods have been replaced by new techniques. The JYD system is a polar system maintained by optical means, and thus it cannot fit in with the needs of establishing a high-precision reference system in modern society. In addition, JYD is not a truly independent polar system and the optical latitude survey data in overseas countries are on the brink of exhaustion. Consequently, this system can barely be maintained. Changing the polar origin to make it consistent with the international system is therefore inevitable.

We also have to make clear that the ellipsoid is oriented to minimize the sum of the squared height anomalies within China by the XAS80 based on the ellipsoidal parameters recommended by IUGG1975, which means that only the three unknowns ΔX_0, ΔY_0, and ΔZ_0 in (7.35) have to be solved. If the ellipsoidal parameters are not fixed in advance, finding solutions based on (7.35) will result in ΔX_0, ΔY_0, ΔZ_0, Δa, and Δe^2, i.e., five parameters. It is evident that such an orientation can allow the ellipsoid surface to be a closer fit to the geoid (or quasi-geoid) in China. As indicated by several computational results over the period 1964–1977, the value of a ranges from 6,378,666 to 6,378,685 m and that of f from 1:291.6 to 1:292.2. It can be seen from the above that the semimajor axis of the calculated ellipsoid based on the regional geoid in China is 500 m longer than that of the obtained ellipsoid based on a global geoid at present. Their denominators of

flattening differ from each other by approximately 6 units due to the complex terrain environment in China. As a result, the ellipsoid parameters based on China's regional geoid have not been adopted in the XAS80.

Highlights of the XAS80 can be summarized as:

1. It is a local coordinate system.
2. It has adopted four ellipsoidal parameters, including both geometric parameters and physical parameters. The numerical values are those recommended at the 16th general assembly of IUGG held in 1975 (see Table 4.1).
3. It employs astro-geodetic positioning and orientation. We select 922 points evenly throughout the country at a $1^{\circ} \times 1^{\circ}$ interval and form an arc measurement equation; the values of ξ_0, η_0, and ζ_0 solved at the geodetic origin are $\xi_0 = -1.9''$, $\eta_0 = -1.6''$, and $\zeta_0 = -14.0$ m.
4. It has a well-defined orientation. The minor axis of the Earth ellipsoid is defined as parallel to the direction from the Earth's center of mass to the polar origin $JYD_{1968.0}$, and the planes of the initial geodetic and astronomical meridians are parallel to each other in China; $\varepsilon_X = \varepsilon_Y = \varepsilon_Z = 0$.
5. The origin is located at Yongle township, a small town 60 km north of Xi'an, Jingyang county, Shaanxi province, in the central area of China, where the precisions of estimated coordinates are relatively homogenous. It can be abbreviated as the Xi'an origin. The estimated values of geodetic longitude and latitude are $L_0 = 108^{\circ}55'$, $B_0 = 34^{\circ}32'$.
6. The coordinates of nearly 50,000 stations are obtained by an integrated adjustment of the nationwide astro-geodetic networks after the establishment of the XAS80.

The XAS80 demonstrates a marked superiority over BJS54. For example, it is in complete conformity with the principles for establishing a classical local coordinate system and can be easily accounted for; the number of the Earth ellipsoid parameters and the magnitude of values are more reliable and accurate; the pointing direction of the coordinate axes is clearly specified; the ellipsoid surface fits the geoid fairly well; and the nationwide mean value of geoid undulations drops from 29 m of BJS54 to 10 m, the maximum value appearing in the southwest of Tibet. For most of the areas in the whole nation the geoid undulation is within 15 m (cf. Fig. 7.9).

In addition, the observational data are reduced rigorously and nationwide unified integrated adjustment is carried out. Thus, the effects caused by the local adjustment and unsuitable network control have been removed, and the accuracy of the adjusted coordinates has been improved.

The changes in ellipsoidal parameters and orientation will inevitably cause variations in geodetic coordinates. Admittedly, the positions of sheet lines will also undergo changes, the degree of which will vary with the point positions. In areas east of 102° east longitude in China, the maximum change is approximately 80 m, with an average of about 60 m. If the positions of sheet lines change, there

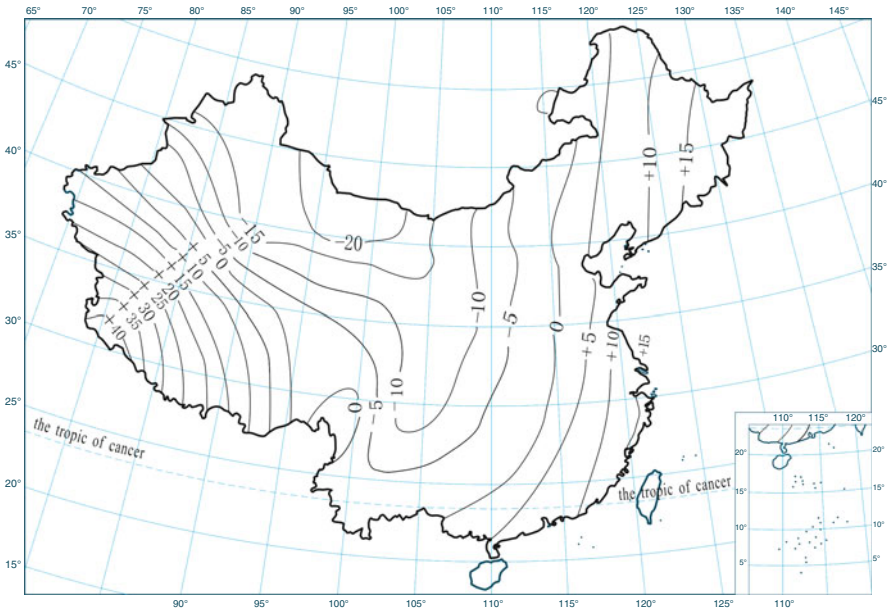


Fig. 7.9 China's mainland geoid, National Geodetic Coordinate System 1980

will be seams on the newly drawn topographic maps due to imperfect splicing. For instance, a change of 80 m will be 1.6 mm on a topographic map at the scale of 1:50,000. The changes in the positions of grid lines are dependent on the changes in geodetic coordinates. The additional effects of the variations in the projected plane coordinates caused by changes in the ellipsoidal parameters are also included. For example, when the geodetic longitude is 116° and geodetic latitude is 46°, the changing values of the grid lines in the *x*- and *y*-directions are 46 m and 55 m, respectively. The changes on the 1:50,000 scale topographic maps are 0.9 mm and 1.1 mm, respectively. The biggest problems generated by implementation of the XAS80 are the changes in positions of the map sheet lines and the grid lines, which have made it troublesome while splicing together the topographic maps based on the old and new geodetic coordinate systems.

7.5.3 Beijing Coordinate System 1954 (New)

The scientific nature, rigor, and advancement of the XAS80 at that time were acknowledged unanimously by everyone, while the results provided by BJS54 were based on local adjustments, manifesting many drawbacks that made it inappropriate for further use. If BJS54 were replaced by XAS80 as the plotting coordinate system (particularly topographic maps at scales larger than 1:50,000),

there would be a great volume of work to connect the old and new coordinate systems. Therefore, implementing a new map coordinate system not only involves its scientific worth and accuracy but also its practicability, feasibility, economic returns, and social benefits. We should not only take into consideration the history of surveying and mapping over the past three decades and its status quo, but also its development in the days and years to come.

The new BJS54 emerged in such a context. The results offered by this coordinate system are obtained based on the XAS80 by transforming the GRS75 Ellipsoid into the Krassowski Ellipsoid and then translating and transforming along the three spatial axes. Hence, its coordinates have not only embodied the superiority of the integrated adjustments, i.e., its accuracy is the same as that of the XAS80, but have also overcome the deficiencies of local adjustment of the original BJS54. In addition, the new BJS54 employs the same ellipsoidal parameters as the old BJS54, and the orientation of the new BJS54 also approximates that of the old BJS54; thus its coordinate values do not differ much from those of the old BJS54 obtained from local adjustment. According to statistics, as for the projected plane coordinates, the coordinate difference between the new and old BJS54 is within 5 m in 80 % of the area of China. Those with a coordinate difference of greater than 5 m are primarily in the northeast and those greater than 10 m are only in a few fringe areas, with the maximum difference being 12.9 m. The difference along the vertical x-axis ranges from -6.5 to +7.8 m, and the difference along the horizontal y-axis ranges from -12.9 to +9.0 m. Such differences are still within the scope of the differences between the provisional coordinates and the adjusted coordinates, reflected on the 1:50,000 scale topographic maps, and most of them are less than 0.1 mm. In this sense, the splicing between the old and new maps will be almost seamless. Hence, the new map not only employs the high-precision integrated adjustment but also brings clear economic benefits because no particular treatment needs to be carried out in order to connect together the old and new maps. Especially in the military system, maps that can show a larger area are mostly at a scale smaller than 1:50,000. Using this new BJS54 as the map coordinate system has some distinct advantages in terms of updating maps, offering fast combat support and making it convenient for officers and men to use the maps.

The relationships between the BJS54, XAS80 and the new BJS54 are depicted vividly in Fig. 7.10.

In Fig. 7.10, $O_{1980}-X_{1980}Y_{1980}Z_{1980}$, $O_{\text{new}1954}-X_{\text{new}1954}Y_{\text{new}1954}Z_{\text{new}1954}$, and $O_{1954}-X_{1954}Y_{1954}Z_{1954}$ are the spatial rectangular coordinate systems with respect to the XAS80, the new BJS54, and the old BJS54, respectively. The direction of the Z_{1954}-axis is not specified. It is defined as non-parallel to the direction from the Earth's center of mass to the polar origin $\text{JYD}_{1968.0}$. The Krassowski Ellipsoid parameters are adopted and the size of the ellipsoid is not drawn in order to keep the map clear. The coordinate systems of the $O_{\text{new}1954}-X_{\text{new}1954}Y_{\text{new}1954}Z_{\text{new}1954}$ and

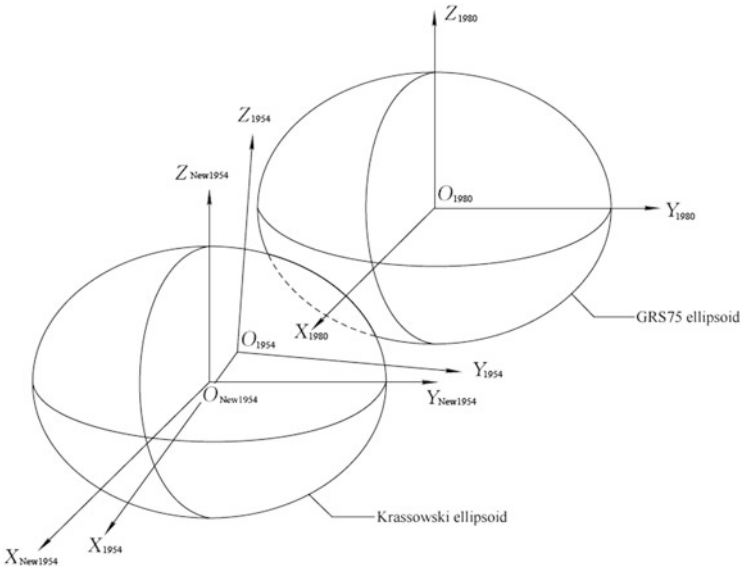


Fig. 7.10 Relationships between three local coordinate systems in China

$O_{1980} - X_{1980}Y_{1980}Z_{1980}$ are parallel to each other. The direction of their Z-axis is parallel to the direction from the Earth's center of mass to the polar origin $\mathrm{JYD}_{1968.0}$ and the X-axis is situated in the plane of the initial geodetic meridian.

It can be seen from Fig. 7.10 that the new BJS54 is a local coordinate system obtained from translating XAS80 in space and replacing the GRS75 Ellipsoid parameters with the Krassowski Ellipsoid parameters. Its translations have the same values as the positioning parameters ΔX_0, ΔY_0, and ΔZ_0 (with positive or negative signs) solved according to (7.35) in XAS80, but of opposite sign.

Hence, the formulae for transformations between the two Cartesian coordinate systems are:

$$\left.\begin{aligned} X_{\mathrm{new1954}} &= X_{1980} - \Delta X_o \\ Y_{\mathrm{new1954}} &= Y_{1980} - \Delta Y_o \\ Z_{\mathrm{new1954}} &= Z_{1980} - \Delta Z_o \end{aligned}\right\}, \tag{7.36}$$

and the formulae for transformations between the two local coordinate systems are:

$$\left.\begin{aligned} L_{\mathrm{new1954}} &= L_{1980} - \mathrm{d}L \\ B_{\mathrm{new1954}} &= B_{1980} - \mathrm{d}B \\ H_{\mathrm{new1954}} &= H_{1980} - \mathrm{d}H \end{aligned}\right\}, \tag{7.37}$$

where

$$\begin{bmatrix} \mathrm{d}L \\ \mathrm{d}B \\ \mathrm{d}H \end{bmatrix} = \begin{bmatrix} -\dfrac{\sin L}{(N+H)\cos B}\rho'' & \dfrac{\cos L}{(N+H)\cos B}\rho'' & 0 \\ -\dfrac{\sin B\cos L}{M+H}\rho'' & -\dfrac{\sin B\sin L}{M+H}\rho'' & \dfrac{\cos B}{M+H}\rho'' \\ \cos B\cos L & \cos B\sin L & \sin B \end{bmatrix}_{1980} \begin{bmatrix} \Delta X_o \\ \Delta Y_o \\ \Delta Z_o \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \dfrac{N}{(M+H)a}e^2\sin B\cos B\rho'' & \dfrac{M(2-e^2\sin^2 B)}{(M+H)(1-f)}\sin B\cos B\rho'' \\ -\dfrac{N}{a}\left(1-e^2\sin^2 B\right) & \dfrac{M}{1-f}\rho''\left(1-e^2\sin^2 B\right)\sin^2 B \end{bmatrix}_{1980} \begin{bmatrix} \Delta a \\ \Delta f \end{bmatrix}. \tag{7.38}$$

We have to point out that, while calculating the orientation elements $\Delta X_0, \Delta Y_0, \Delta Z_0$ based on $\Sigma_1^{922}\xi_{1980}^2 = \text{minimal solution}$ according to (7.35), the Euler angles generated by the inconsistency of the axis directions between XAS80 and BJS54 are not taken into consideration. Consequently, the values of ΔX_0, ΔY_0, and ΔZ_0, in the strict sense, are not the three components of the rectangular coordinates between the ellipsoidal centers O_{1980} and O_{1954} of the above two geodetic coordinate systems. Hence, when the values of ΔX_0, ΔY_0, and ΔZ_0 of the opposite signs are transformed into the new BJS54, the ellipsoidal center is $O_{\text{new}1954}$ instead of O_{1954}, namely the ellipsoidal center of the new BJS54 does not coincide with that of the BJS54. The difference i, however, very slight.

The geodetic origin of the XAS80 is located in Xi'an. The integrated adjustment of the nationwide geodetic networks has been carried out based on the geodetic origin data. In addition, the coordinates in the system of the new BJS54 have been transformed from the XAS80, and thus the origin of the new BJS54 still remains the Xi'an origin. However, the geodetic origin data of the two local coordinate systems are different, which can be calculated according to (7.38). The result shows that their geodetic longitudes differ from each other by $-2.19''$, while geodetic latitudes differ by $0.39''$.

Highlights of the new BJS54 can be summarized as follows:

1. It is a local coordinate system.
2. It has adopted the Krassowski Ellipsoid parameters (see Table 4.1).
3. It employs the astro-geodetic orientation. The ellipsoidal center approximates very well that of the BJS54, but they are not coincident with each other.
4. It has well-defined orientation. The minor axis of the Earth ellipsoid is defined as parallel to the direction from the Earth's center of mass to the polar origin $\text{JYD}_{1968.0}$, and the planes of the initial geodetic and astronomical meridians are parallel to each other in China, thus $\varepsilon_X = \varepsilon_Y = \varepsilon_Z = 0$.

5. The origin is located at Yongle township, Jingyang county, Shaanxi province but with different geodetic origin data from the XAS80.
6. Coordinates obtained are transformed from the XAS80 after the integrated adjustment. The precision of the coordinates is exactly the same as for the XAS80.
7. Used as a plotting coordinate system, for maps at a scale smaller than 1:50,000, the new and old maps splice (the break of edge-matching between the new and old maps) together almost seamlessly.

The advantages of the new BJS54 lie in the fact that it has overcome the low accuracy of the adjusted coordinates of the old BJS54 from the local adjustments, and the newly drawn topographic maps based on the new BJS54 and the old topographic maps splice together well. Generally, for maps at a scale smaller than 1:50,000, there will be no apparent seams. However, for large-scale maps of 1:25,000 or 1:10,000, seams will arise during splicing. In terms of demand for production and scientific development, digital products will replace the simulation products. Hence, the new BJS54 has not been practically used in the field of analog mapping.

7.5.4 Geocentric Coordinate System 1978

The geocentric coordinate system 1978 was obtained by transforming the BJS54 based on DX-1 coordinate transformation parameters.

DX-1 has only three translation parameters, i.e., $\Delta X_0, \Delta Y_0, \Delta Z_0$. The transformation parameters of DX-1 were defined and established in some relevant conferences in November 1978. At that time, satellite geodetic techniques were still in their infancy, so the geocentric coordinates could not be observed directly. DX-1 has been established by employing five methods, as identified below:

1. Astro-gravimetric method
2. Global astro-geodetic geoid undulations
3. Difference between astro-geodetic and gravimetric geoid undulations
4. Using Doppler receivers like MX-702A to determine the transit navigation satellite system (TRANSIT System) to establish geocentric coordinates
5. Using Doppler receivers like CMA-722B to determine the transit navigation satellite system (TRANSIT System) to establish geocentric coordinates

The translation parameters can be obtained by any one of the above methods alone. If different weights are assigned to these five methods, respectively, taking the weighted mean will produce the three translation parameters of the DX-1, ΔX_0, ΔY_0, and ΔZ_0.

The three translation parameters ΔX_0, ΔY_0, and ΔZ_0 denote the displacements of the center position of the BJS54 reference ellipsoid deviating from the geocenter, namely the three coordinate components of the center of the BJS54 in the

geocentric coordinate system. The obtained system based on this set of parameters is named the Geocentric Coordinate System 1978.

Based on the DX-1 transformation parameters, we can transform the geodetic Cartesian coordinates in BJS54 into geocentric Cartesian coordinates in the Geocentric Coordinate System 1978:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{1978} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{1954} + \begin{bmatrix} \Delta X_o \\ \Delta Y_o \\ \Delta Z_o \end{bmatrix}_{\text{DX-1}} . \tag{7.39}$$

DX-1 excludes rotation and scale parameters. It was only a preliminary result to meet the urgent needs of space technologies back then. Shortcomings were unavoidable in the course of its establishment. For instance, the materials and data used in calculations were limited and inaccurate, and the data processing methods were not optimal. Hence, the direction of the coordinate axis of the geocentric coordinates obtained by DX-1 transformation parameters is not clearly defined. It is estimated that the mean square error of the coordinate component of the Geocentric Coordinate System 1978 is approximately ± 10 m.

7.5.5 Geocentric Coordinate System 1988

Since 1979, some departments have done plenty of work in developing space geodetic techniques. For instance, China established a Doppler network with 37 points across the country in 1980; another dynamic satellite geodetic network was established in 1982, so geocentric coordinates with high precision were obtained. In the meantime, an integrated adjustment of the nationwide astronomical geodetic network was carried out, by which the geodetic coordinates of nearly 50,000 stations were determined. The height anomalies nationwide have been obtained with quite high precision by means of the astro-gravimetric leveling. All kinds of geocentric coordinate systems are continually updated internationally, and some quite new data have appeared. All have created conditions for China to refine its geocentric coordinate transformation parameters.

After years of preparation, some departments concerned established the DX-2 steering group for integrated data processing in May 1987. Eventually, by the end of 1988, the geocentric coordinate transformation parameters of DX-2 were completed.

The transformation parameters of DX-2 are obtained by integrating three methods for establishing geocentric coordinate parameters:

1. Geocentric coordinates obtained by using an MX-1502 Doppler receiver to the track transit satellite system (TRANSIT System) (Doppler network with 37 points nationwide)

2. Geocentric coordinates obtained by dynamic satellite geodesy (seven points nationwide)
3. Global astro-geodetic geoid undulations (referenced to the XAS80)

DX-2 consists of seven transformation parameters. The XAS80 coordinates and the new BJS54 coordinates can both be transformed into geocentric coordinates based on DX-2 transformation parameters, and the corresponding system is called the Geocentric Coordinate System 1988. Thus, DX-2 has two sets of transformation parameters, i.e., DX-2_{new1954} and DX-2_{1980}; the geocentric coordinates obtained are completely coincident.

The origin of the Geocentric Coordinate System 1988 is the Earth's center of mass. The Z-axis is directed towards the CIO (BIH1968); the X-axis is directed towards the international zero-longitude (BIH1968); the Y-axis, Z-axis, and X-axis constitute a right-handed coordinate system. The unit of length is the meter. If represented by geodetic coordinates, the ellipsoidal parameters are those of GRS75, as shown in Table 4.1.

When the Cartesian coordinates in XAS80 or in the new BJX54 are transformed into the coordinates in the Geocentric Coordinate System 1988, the formulae for computation are:

$$\left.\begin{aligned} X_D &= X(1+m) + Y\cdot\varepsilon_Z''/\rho'' - Z\cdot\varepsilon_X''/\rho'' + \Delta X_0 \\ Y_D &= Y(1+m) - X\cdot\varepsilon_Z''/\rho'' + Z\cdot\varepsilon_X''/\rho'' + \Delta Y_0 \\ Z_D &= Z(1+m) + X\cdot\varepsilon_Y''/\rho'' - Y\cdot\varepsilon_X''/\rho'' + \Delta Z_0 \end{aligned}\right\}, \tag{7.40}$$

where X_D, Y_D, and Z_D denote the coordinates in the Geocentric Coordinate System 1988, and X,Y, and Z represent the coordinates in XAS80 or in the new BJS54, respectively. Their corresponding transformation parameters are DX-2_{1980} or DX-2_{new1954}.

It is estimated that the mean square error of each component of the geocentric coordinate obtained from DX-2 is better than ± 5 m.

7.5.6 China Geodetic Coordinate System 2000

Since the 1980s, under the concerted efforts of the Chinese State Bureau of Surveying and Mapping (SBSM), the Surveying and Mapping Bureau of the General Staff Headquarters (SMBGSH), China Seismological Bureau (CSB), Chinese Academy of Sciences (CAS), and some other units, China has established in succession first- and second-order nationwide GPS networks, order A and B national GPS networks, GPS continuously operating reference station (CORS) networks, and Crustal Movement Observational Network of China (CMONOC). A large-scale combined adjustment between terrestrial and spatial networks has been carried out. These achievements have marked the mature technological

conditions for China to construct its own new geocentric coordinate frames (see Wei 2003; Ning 2002).

Definition of the CGCS2000

The new generation national geodetic coordinate system in China is the China Geodetic Coordinate System 2000, abbreviated as CGCS2000 (see Chen 2008; Yang 2009). Its definition is consistent with that of the CTRS, namely:

1. Origin: coincides with the center of mass of the whole Earth including oceans and atmosphere
2. Orientation: initially given by the BIH orientation at 1984.0
3. Time evolution (of the orientation): to create no residual global rotation relative to the crust
4. Unit of length: the international standard meter, defined in a local Earth frame in the meaning of a relativistic theory of gravitation

The reference epoch of CGCS2000 is at 2000.0.

The reference ellipsoid of CGCS2000 is the CGCS2000 Reference Ellipsoid, whose defined parameters are shown in Table 4.1. The normal ellipsoid is consistent with the reference ellipsoid.

Several points need to be addressed:

1. Of the four constants of the CGCS2000 Reference Ellipsoid, three constants are coincident with GRS80, namely a, f, and ω. GM is the value recommended by IERS, and most of the other geocentric ellipsoids in the world are the same.
2. There is a slight difference between the values a and f of CGCS2000 and the higher-precision values recommended by IERS ($a = 6,378,136.6$ m, $1/f = 298.25642$), but this will have no effect on applications.
3. The CGCS2000 Reference Ellipsoid differs slightly from WGS84 only in the flattening f (which generated 1 mm error of the terrestrial coordinates at the equator). These two ellipsoids can therefore be considered practically coincident with each other.

Realization of the CGCS2000

The CGCS2000 is realized by estimates of the station coordinates and velocities at three levels.

Level 1: Continuously Operating Reference System (CORS). These constitute the fundamental frame of CGCS2000 with an accuracy of a few millimeters in coordinates and 1 mm per annum in velocity. The CORS also provide coordinate datum for static and dynamic positioning and navigation.

GPS CORS are permanent ground-based observation stations that are designed to provide long-term and continuous tracking and receiving of GPS

satellite signals. They provide users of different kinds with different needs and at different levels automatically with verified GPS observational data (carrier phrase, pseudorange), all kinds of corrections, GPS status messages and other GPS-relevant services, and programs in real-time by using data communications and internet technologies. Currently, the CGCS2000 is maintained by 260 CORS stations and its coordinate accuracy is at the level of a millimeter.

Level 2: GPS geodetic control network (see Yang et al. 2009). This consists of all the nationwide GPS networks including orders A and B national GPS networks, the first- and second-order nationwide GPS networks, the Crustal Movement Observation Network of China, and all the other GPS networks that monitor crustal deformations except CORS. Its three-dimensional geocentric coordinates are at centimeter-level accuracy, and velocity precision is 2–3 mm per annum.

The order A and order B GPS networks were observed by the SBSM between 1991 and 1997. The order A network consists of 30 principal points and 22 auxiliary points. The order B network comprises 818 points. The order A and B network adjustments are referred to ITRF93 at the epoch of 1996.365. The geocentric coordinates of the points after adjustment have an accuracy of 10^{-7} order of magnitude.

The first- and second-order national GPS networks were both observed by SMBGSH in China over the years 1991–1997. The first-order network consists of 44 points and the second-order network of 534 points. The first- and second-order network adjustments are referred to ITRF96 at the epoch of 1997.0. The geocentric coordinates of the points after adjustment have an accuracy of 10^{-8} order of magnitude.

The Crustal Movement Observation Network of China was jointly observed by the CSB in collaboration with SMBGSH, CAS, and SBSM in China from 1998 to 2002. It includes datum networks, basic networks, and local networks. There are 29 datum points, 56 basic points, and 1,000 local points. The network adjustment is referred to ITRF96 at the epoch of 1998.680. The geocentric coordinates of the points after adjustment are with a general accuracy of better than 10^{-8} order of magnitude.

The GPS geodetic networks together with CORS constitute the CGCS2000 frame.

Level 3: The astro-geodetic network (see Yang et al. 2005). This consists of approximately 50,000 astro-geodetic points after the combined adjustment between terrestrial and spatial networks. The error of the geodetic longitude and latitude is within 0.3 m and the error of the geodetic height does not exceed 0.5 m.

GPS geodetic networks in China are at lower density than astro-geodetic networks, only approximately one-twentieth of the latter. Hence, the low-density 3D geocentric coordinate reference frame offered by GPS geodetic networks cannot fully realize the 3D geocentric coordinate reference system for engineering purposes. Combined adjustments between the national astro-geodetic networks and GPS geodetic control networks were carried out in China by the SBSM and SMBGSH, respectively. The high-precision geocentric

coordinates of 48,919 first- and second-order astro-geodetic points were obtained with an average accuracy of position of ±0.11 m, which bettered the precision and current situation of the national astro-geodetic networks and improved the density and distribution of the CGCS2000 reference frame.

Difference Between CGCS2000 and ITRF

By definition, CGCS2000 and ITRF belong to the same coordinate system. The coordinates of CGCS2000 are determined by GPS relative positioning techniques, while the datum station coordinates are in the ITRF reference frame. Hence, CGCS2000 and ITRF belong to the same reference frame. With proper observations and reasonable calculations, we can strive to keep CGCS2000 consistent with ITRF within 2 cm. Hence, CGCS2000 is the Chinese equivalent to ITRF but of a higher density. The difference between CGCS2000 and ITRF is negligible when the accuracy is at the level of a centimeter.

Difference Between CGCS2000 and the Current Local Coordinate System

Different Methods for Ellipsoid Orientation

A local coordinate system is designed to establish the ellipsoid that best fits the geoid over the local area by means of an astro-geodetic orientation, whose origin deviates considerably from the geocenter (approximately 200 m). Taking XAS80 as an example, its reference ellipsoid fits the geoid well on a national basis. Two isolines of zero height anomalies pass through the east and west of China. The height anomalies in most areas are within 20 m. The effect of the height anomalies on distance is smaller than 1/300,000. CGCS2000 is a geocentric coordinate system, which is realized by modern geodetic techniques based on the techniques employed in the establishment of a terrestrial reference frame. The center of the ellipsoid defined by CGCS2000 is coincident with the Earth's center of mass. Meanwhile, the ellipsoid also most closely fits the geoid on a global basis. The ellipsoids that are global "best fits" to the geoid may not necessarily be local "best fits" to the geoid.

Different Realization Techniques

The local coordinate system is established by classical geodetic techniques, whereas CGCS2000 is established by space geodetic techniques.

Different Coordinate System Dimensions

The local coordinate system is a 2D coordinate system, whereas CGCS2000 is a 3D coordinate system.

Different Coordinate System Origins

The origin of the local coordinate system deviates considerably from the Earth's center of mass, whereas the origin of CGCS2000 is coincident with the Earth's center of mass.

Different Precisions of Realization

The local coordinate system lacks high-precision external control, so the long-distance precision is quite low. The accuracy of CGCS2000 has increased tenfold compared to that of the current ellipsoid centric coordinate system. Its relative precision can be up to 10^{-7}–10^{-8}.

Impact and Countermeasures of Implementing CGCS2000

The geodetic coordinate system is fundamental to the all surveying and mapping work. The change or a replacement of a coordinate system involves a many aspects, but the most immediate consequences will be regarding utilization of the surveying and mapping data, results, and products under the old coordinate system. This mainly touches upon the following aspects:

1. Results of geodetic points under the old coordinate system (approximately over 200,000 points)
2. Topographic maps, nautical charts, aerocharts, and cadastral maps under the old coordinate system
3. Results of urban independent coordinate systems (including coordinates of the control points and topographic maps)
4. The established geographical spatial data infrastructure and geographic information system
5. Gravity anomalies, height anomalies, and deflections of the vertical of other ellipsoids
6. Surveying and mapping results for special applications

For issues concerning the transitional applications of geodetic results and mapping products, the following technical countermeasures need to be taken:

1. We carry out a combined adjustment between terrestrial and spatial networks and incorporate the geodetic points of various-order networks into the reference

system of CGCS2000. The geodetic points for which a combined adjustment cannot be carried out should be corrected to CGCS2000 by precise coordinate transformations.

2. We obtain precisely the transformation parameters from BJS54, XAS80, and urban independent coordinate systems to CGCS2000 and compile corresponding coordinate transformation software dedicated to public use. The accuracy indicators of the coordinate transformation can be sorted into the following three categories: for low-precision transformation in common navigation applications, the mean square error of each coordinate component is in the range of 5–10 m; for medium precision transformation as a 3D coordinate similarity transformation, the mean square error of each coordinate component is in the range of 0.5–5 m; for high-precision transformation in sheet lines of topographic maps and coordinate grid transformations, the mean square error of each coordinate component should not exceed 0.5 m.
3. The printed paper topographic maps in the old coordinate system can be corrected to the CGCS2000 by replotting the neat lines and grid lines. Computation shows that, while changing the coordinate system for the maps at scales larger than 1:1,000,000, the changing value in the geographical location of the points on the maps will exceed the mapping accuracy, and the maps should be replotted. However, when the distance and azimuth shift between any pair of points on a map at any scale are within the mapping accuracy, the changes can be neglected. Hence, for the correction of existing paper maps, all that needs to be done is to translate the sheet lines and the square grids of each map.
4. For digital topographic maps, we transform the coordinates of all the map layers and map elements using software, followed by cutting and clipping over again.
5. The gravity anomaly that is referenced to other ellipsoids can be transformed to the gravity anomaly referred to the CGCS2000 ellipsoid using software based on the normal gravity formula. Likewise, the height anomalies and deflections of the vertical that are referenced to other ellipsoids can also be converted to that referred to CGCS2000.

It is perfectly viable to ensure a smooth transition from the old to the new coordinate system through a well-conceived plan and well-arranged implementation by using the existing technologies available in China. The CGCS2000 was put into effect on July 1st, 2008 in China. With regards to the transformation and connection between CGCS2000 and the existing national geodetic coordinate systems (referring to BJS54 and XAS80), the transition period is regulated to take 8–10 years according to the correlative provisions.

Review and Study Questions

1. Draw a geometric figure to explain the concept of Euler angles.
2. By referring to the “double-parallel conditions” (i.e., the minor axis of the ellipsoid is parallel to the Earth’s rotation axis, and the planes of the initial

geodetic and astronomical meridians are parallel to each other), explain the geometric meaning of β_2, γ_3, and provide the relationship between β_2, γ_3, and Euler angles.

3. Draw a geometric figure to derive the Bursa–Wolf model for transformations between different geodetic Cartesian coordinate systems.
4. Explain how to solve the transformation parameters of different geodetic Cartesian coordinate systems using the Bursa–Wolf model.
5. What is the geodetic origin data and what is its relation to the ellipsoid orientation?
6. What are the conditions of the ellipsoid orientation and how could these conditions be satisfied?
7. Explain the concepts of a single astronomical position datum orientation and the astronomical–geodetic orientation.
8. Account for the following concepts: the Earth's polar origin, the origin of longitude, the polar coordinate system, the ideal TRS, the CTRS, and the conventional terrestrial reference frame.
9. Illustrate the datum definition adopted by ITRF*yy* (where *yy* is the annual series of this frame) at different dates.

Bibliography

Abbondanza, C., Altamimi, Z., Sarti, P., Negusini, M., & Vittuari, L. (2009). Local effects of redundant terrestrial and GPS-based tie vectors in ITRF-like combinations. *Journal of Geodesy, 83*(11), 1031–1040. doi:10.1007/s00190-009-0321-6.

Altamimi, Z. (2009). ETRS89 realization: current status, ETRF2005 and future development. *Bulletin of Geodesy and Geomatics, LXVIII*, 3.

Altamimi, Z., Boucher, C., & Gambis, D. (2005a). Long-term stability of the terrestrial reference frame. *Advances in Space Research, 36*, 342–349. doi:10.1016/j.asr.2005.03.068.

Altamimi, Z., Boucher, C., & Sillard, P. (2002a). New trends for the realization of the international terrestrial reference system. *Advances in Space Research, 30*(2), 175–184. doi:10.1016/S0273-1177(02)00282-X.

Altamimi, Z., Boucher, C., & Willis, P. (2005b). Terrestrial reference frame requirements within GGOS perspective. *Journal of Geodynamics, 40*, 363–374. doi:10.1016/J.jog.2005.06.002.

Altamimi, Z., & Collilieux, X. (2009). IGS contribution to ITRF. *Journal of Geodesy, 83*(3/4), 375–383. doi:10.1007/s00190-008-0294-x.

Altamimi, Z., & Collilieux, X. (2010). Quality assessment of the IDS contribution to ITRF2008. *Advances in Space Research, 45*(12), 1500–1509. doi:10.1016/j.asr.2010.03.010.

Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B., & Boucher, C. (2007). ITRF2005: A new release of the international terrestrial reference frame based on time series of station positions and Earth orientation parameters. *Journal of Geophysical Research, 112*, B09401. doi:10.1029/2007JB004949.

Altamimi, Z., Collilieux, X., & Métivier, L. (2011). ITRF2008: an improved solution of the international terrestrial reference frame. *Journal of Geodesy, 85*(8), 457–473. doi:10.1007/s00190-011-0444-4.

Altamimi, Z., Collilieux, X., & Métivier, L. (2012b). ITRF combination: theoretical and practical considerations and lessons from ITRF2008. In *Proceedings of the IAG Symposium. REFAG2010. Marne-La-Vallée* (International Association of Geodesy Symposia, Vol. 138). Springer, doi:10.1007/978-3-642-32998-2_2.

Altamimi, Z., & Dermanis, A. (2012). The choice of reference system in ITRF formulation. In *VII Hotine-Marussi Symposium on Mathematical Geodesy: Proceedings of the Symposium in Rome, 6–10 June, 2009* (International Association of Geodesy Symposia, Vol. 137, pp. 329–334). doi: 10.1007/978-3-642-22078-4_49.

Altamimi, Z., Métivier, L., & Collilieux, X. (2012c). ITRF2008 plate motion model. *Journal of Geophysical Research, 117*, B07402. doi:10.1029/2011JB008930.

Altamimi, Z., Sillard, P., & Boucher, C. (2002b). ITRF2000: A new release of the international terrestrial reference frame for Earth science applications. *Journal of Geophysical Research, 107*(B10), 2214. doi:10.1029/2001JB000561.

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5,

Altamimi, Z., Sillard, P., & Boucher, C. (2003). The impact of a no-net-rotation condition on ITRF2000. *Geophysical Research Letters, 30*(2), 1064. doi:10.1029/2002GL016279.

Altamimi, Z., Angermann, D., Argus, D., Blewitt, G., Boucher, C., Chao, B., et al. (2001). The terrestrial reference frame and the dynamic Earth. *EOS Transactions American Geophysical Union, 82*(25):273–279, Washington, D.C. doi:10.1029/EO082i025P00273-01.

Angermann, D. (2012). Standards and conventions for geodesy. In H. Drewes, H. Hornik, J. Adam, & S. Rozsa (Eds.), The geodesist's handbook 2012. *Journal of Geodesy*, *86*(10), 961–964. doi:10.1007/s00190-012-0584-1.

Angermann, D., Drewes, H., Gerstl, M., Meisel, B., Seitz, M., & Thaller, D. (2010a). GGOS-D global terrestrial reference frame. In F. Flechtner, T. Gruber, A. Güntner, M. Mandea, M. Rothacher, T. Schöne, & J. Wickert (Eds.), *System earth via geodetic-geophysical space techniques* (pp. 555–564). New York: Springer. doi:10.1007/978-3-642-10228-8_46.

Angermann, D., Drewes, H., & Seitz, M. (2012a). Global terrestrial reference frame realization within the GGOS-D project. In S. Kenyon, M. C. Pacino, & U. Marti (Eds.), *Geodesy for planet earth* (IAG Symposia, Vol. 136, pp. 87–93). New York: Springer. doi:10.1007/978-3-642-20338-1_11.

Angermann, D., & Müller, H. (2009). On the strength of SLR observations to realize the scale and origin of the terrestrial reference system. In M. G. Sideris (Ed.), *Observing our changing earth* (IAG Symposia, Vol. 133, pp. 21–27). doi:10.1007/978-3-540-85426-5_3.

Angermann, D., Seitz, M., & Drewes, H. (2012b). Global terrestrial reference systems and their realizations. In G. Xu (Ed.), *Sciences of geodesy—II* (pp. 97–132). Berlin: Springer. doi:10.1007/978-3-642-28000-9_3.

Blewitt, G. (2003). Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth. *Journal of Geophysical Research, 108*(B2), 2203. doi:10.1029/2002JB002082.

Blewitt, G., Altamimi, Z., Davis, J. L., Gross, R. S., Kuo, C., Lemoine, F. G., et al. (2010). *Geodetic observations and global reference frame contributions to understanding sea-level rise and variability. A World Climate Research Programme Workshop and a WCRP contribution to the Global Earth Observation System of Systems* (pp. 256–284). Chichester: Wiley-Blackwell.

Böckmann, S., Artz, T., & Nothnagel, A. (2010). VLBI terrestrial reference frame contributions to ITRF2008. *Journal of Geodesy, 84*(3), 20–209.

Bomford, A. G. (1962). Transverse Mercator arc-to-chord and finite distance scale factor formulae. *Empire Survey Review, 16*(125), 318–327.

Bomford, G. (1980). *Geodesy* (4th ed.). Oxford: Clarendon Press. 731pp. ISBN 019851946X.

Boucher, C., Altamimi, Z., & Duhem, L. (1992). *IERS technical note 12: ITRF91 and its associated velocity field*. Paris: Central Bureau of IERS.

Boucher, C., Altamimi, Z., & Sillard, P. (1999). *The 1997 International terrestrial reference frame (ITRF97)* (IERS Technical Note 27). Paris: Observatoire de Paris.

Boucher, C., Altamimi, Z., Sillard, P., & Feissel-Vernier, M. (2004). *The ITRF2000* (IERS Technical Note 31). Paris: Observatoire de Paris.

Bouman, J., Koop, R., Tscherning, C., & Visser, P. (2004). Calibration of GOCE SGG data using high-low SST—Terrestrial gravity data and global gravity field models. *Journal of Geodesy, 78*, 124–137.

Bruyninx, C, Altamimi, Z., Becker, M., Craymer, M., Combrinck, L., Combrink, A., et al. (2010). A dense global velocity field based on GNSS observations : Preliminary results. In *Proceedings of the IAG 2009 General Assembly*, Buenos Aires.

Bruyninx, C., Altamimi, Z., Boucher, C., Brockmann, E., Caporali, A., Gürtner, W., et al. (2009). *The European reference frame: Maintenance and products, geodetic reference frames,* Munich, Germany, 9–14 October 2006, International Association of Geodesy Symposia (IAG Symposium, Vol. 134, pp. 131–136). Springer. doi:10.1007/978-3-642-00860-3_20.

Casey, J. (2005). *A treatise on spherical trigonometry and its application to geodesy and astronomy with numerous examples* (Michigan Historical Reprint Series). Scholarly Publishing Office, University of Michigan Library. ISBN 9781418180478.

Chen, J. (1988). *Foundation of terrestrial reference system orientation theory*. Beijing: Surveying and Mapping Press. 148pp (in Chinese). ISBN 7503001283.

Chen, J. (1991). *Modern orientation theory of terrestrial reference system and Earth rotation movement*. Beijing: Surveying and Mapping Press. 313pp (in Chinese). ISBN 7503000600.

Chen, J. (2003). Considerations over establishing China modern geodetic datum. *Journal of Wuhan University* (Information Science Edition), 2003, 28(S1): 1–6 (In Chinese).

Chen, J. (2008). Chinese modern geodetic datum—Chinese geodetic coordinate system 2000 (CGCS 2000) and its frame. *Acta Geodaetica et Cartographica Sinica, 37*(3), 269–271 (in Chinese).

Chen, J., Yang, Y., & Wang, M. (2007a). Establishment of China geodetic control network 2000 and its technological progress. *Journal of Surveying and Mapping, 36*(1), 1–8 (in Chinese).

Chen, J., Yang, Y., Wang, M., Zhang, Y., Tang, Y., Li, H., et al. (2007b). Establishment of China 2000 national geodetic control network and its technological progress. *Acta Geodaetica et Cartographica Sinica, 36*(1), 1–8 (in Chinese).

Cheng, P., Cheng, Y., Wen, H., Huang, J., Wang, H., Wang, G. (2008). *Practical manual on CGCS2000*. Beijing: Surveying and Mapping Press (in Chinese).

Clarke, A. R. (1880). *Geodesy*. Oxford: Clarendon Press.

Collier, P. A. (2002). *Development of Australia's national GDA94 transformation grid* (p 8). Consultant's Report to the Intergovernmental Committee on Surveying and Mapping.

Collilieux, X., & Altamimi, Z. (2012). *External evaluation of the origin and the scale of the International terrestrial reference frame, proceedings of the IAG symposium. REFAG2010. Marne-La-Vallée* (International Association of Geodesy Symposia, Vol. 138). Berlin: Springer. doi:10.1007/978-3-642-32998-2_5.

Collilieux, X., Altamimi, Z., Argus, D. F., Boucher, C., Dermanis, A., Haines, B. J., et al. (2012). External evaluation of the terrestrial reference frame: Report of the task force of the IAG sub-commission 1.2. In *Proceedings of the XXV IUGG General Assembly*.

Collilieux, X., Altamimi, Z., Coulot, D., Ray, J., & Sillard, P. (2007). Comparison of very long baseline interferometry, GPS, and satellite laser ranging height residuals from ITRF2005 using spectral and correlation methods. *Journal of Geophysical Research, 112*, B12403. doi:10.1029/2007JB004933.

Collilieux, X., Altamimi, Z., Coulot, D., van Dam, T., & Ray, J. (2010). Impact of loading effects on determination of the international terrestrial reference frame. *Advances in Space Research, 45*, 144–154. doi:10.1016/j.asr.2009.08.024.

Collilieux, X., Métivier, L., Altamimi, Z., van Dam, T., & Ray, J. (2011). Quality assessment of GPS reprocessed terrestrial reference frame. *GPS Solutions, 15*(3), 219–231. doi:10.1007/s10291-010-0184-6.

Cooper, M. A. R. (1987). *Control surveys in civil engineering*. New York: Nichols Publishing Company.

Deakin, R. E., Hunter, M. N., & Karney, C. F. F. (2011). The Gauss-Krueger projection: Karney-Krueger equations. In *Proceedings of the XXV International Cartographic Conference (ICC2011)*, Paris.

DeMers, M. N. (2005). *Fundamentals of geographic information systems* (3rd ed.). Hoboken, NJ: Wiley. ISBN 9814126195.

Denker, H., Barriot, J.-P., Barzaghi, R., Forsberg, R., Ihde, J., Kenyeres, A., et al.. (2005). Status of European gravity and geoid project EGGP. In C. Jekeli, L. Bastos, J. Fernandes (Eds.), Gravity, geoid and space missions—GGSM2004. *IAG International Symposium*, Porto, Portugal, 2004 (IAG Symposia, Vol. 129, pp. 125–130).

Denker, H., Torge, W., Wenzel, G., Ihde, J., & Schirmer, U. (2000) Investigation of different methods for the combination of gravity and GPS/levelling data. In K. P. Schwarz (Ed.),

Geodesy beyond 2000—the challenges of the first decade (IAG Symposia, Vol. 121, pp. 137–142).

Dennis, D. M., & Petit, G. (Eds.). (2004). *IERS conventions (2003)* (IERS Technical Note No. 32). Frankfurt: Verlag des Bundesamts für Kartographie und Geodäsie, 127pp. ISBN 3-89888-884-3.

Dettmering, D., Schmidt, M., Heinkelmann, R., & Seitz, M. (2011). Combination of different space-geodetic observations for regional ionosphere modeling. *Journal of Geodesy, 85*(12), 989–998. doi:10.1007/s00190-010-0423-1.

Dewhurst, W. T. (1990). *NADCON-The application of minimum curvature derived surfaces in the transformation of positional data from the north American datum of 1927 to the north American datum of 1983*. (NOAA Technical Memorandum NOS NGS-50, p. 30).

DMA. (1984). *Geodesy for the layman* (US Defense Mapping Agency Technical Report 80-003). Washington, DC: DMA, 96pp.

Dow, J. M., Neilan, R. E., & Rizos, C. (2009). The International GNSS service in a changing landscape of global navigation satellite systems. *Journal of Geodesy, 83*(3/4), 191–198.

Drewes, H. (Ed.). (2009). *Geodetic reference frames* (IAG Symposia, Vol. 134). Berlin: Springer.

Duan, D. (1996). *Electronic tacheometric technology*. Beijing: Chinese People's Liberation Army Publishing House (in Chinese).

Featherstone, W. E. (1996). An updated explanation of the geocentric datum of Australia and its effects upon future mapping. *The Australian Surveyor, 41*(2), 121–130.

Featherstone, W. E., & Claessens, S. J. (2008). Closed-form transformation between geodetic and ellipsoidal coordinates. *Studia Geophysica et Geodaetica, 52*, 1–18.

Featherstone, W. E., & Rüeger, J. M. (2002). The importance of using deflections of the vertical for reduction of survey data to a geocentric datum. *The Australian Surveyor, 47*(1).

FGDC. (2001). *United States National Grid*. Reston, VA: US Federal Geographic Data Committee FGDC.

Flechtner, F., Gruber, T., Guntner, A., Mandea M., Rothacher M., Schöne T., et al. (Eds.). (2010). *System earth via geodetic-geophysical space techniques*. Berlin: Springer.

Galati, S. R. (2006). *Geographic information systems demystified*. London: Artech House. ISBN 158053533X.

Gendt, G., Altamimi, Z., Dach, R., Söhne, W., & Springer, T. (2010). GGSP: Realization and maintenance of the Galileo terrestrial reference frame. *Advances in Space Research, 47*(2), 174–185. doi:10.1016/j.asr.2010.02.001.

Gitlein, O., Denker, H., & Müller, J. (2005) Local geoid computation by the spectral combination method, In C. Jekeli, L. Bastos, J. Fernandes (Eds.), *Gravity, geoid and space missions—GGSM2004* (IAG Symposia, Vol. 129, pp. 179–184).

Grewal, M. S., Weill, L. R., & Andrews, A. P. (2001). *Global positioning systems, inertial navigation, and integration*. New York: Wiley. 392pp. ISBN 0471-35032-X.

Gross, R. S. (2007). Earth rotation variation-long period. In T. A. Herring (Ed.), *Physical geodesy, treatise on geophysics* (Vol. 3). Amsterdam: Elsevier.

Gurtner, W., Noomen, R., & Pearlman, M. R. (2005). The international laser ranging service: current status and future developments. *Advances in Space Research, 36*(3), 327–332.

Hagedoorn, J. M., Wolf, D., & Martinec, Z. (2007). An estimate of global mean sea-level inferred from tide-gauge measurements using glacial-isostatic models consistent with the relative sea-level record. *Pure and Applied Geophysics, 164*, 791–818.

Harvey, B. R. (1986). Transformation of 3-D coordinates. *The Australian Surveyor, 32*(2), 105–125.

Heinkelmann, R., Seitz, M., Mora-Diaz, J., Bloßfeld, M., Gerstl, M., & Schmid, R. (2013). *DGFI Analysis Center annual report 2012* (IVS Annual Report 2012).

Helmert, F. R. (1880). *Die mathematischen und physikalischen theorieen der höheren geodäsie* (Vol. 1). Leipzig: Die mathematischen Theorieen.

Hofmann, W. B., & Moritz, H. (2005). *Physical geodesy*. Wien: Springer. 403pp. ISBN 3211235841.

Hooijberg, M. (2008). *Geometrical geodesy—Using information and computer technology*. Berlin: Springer.

Hotine, M. (1946, 1948). The orthomorphic projection of the spheroid, parts I–V. *Survey Review*, *8* (62), 301–311 and *9*(63), 29–35; *9*(64), 52–70; *9*(65), 112–123; *9*(66), 157–166.

Hristow, W. K. (1943). *Die Gauss-krüger'schen koordinaten auf dem ellipsoid*. Leipzig: Verlag B. G. Teubner. 254pp.

Hristow, W. K. (1955). *Die Gaussschen und geographischen koordinaten auf dem ellipsoid von Krassowsky*. Berlin: Verlag Technik.

Hugentobler, U., Gruber, T., Steigenberger, P., Angermann, D., Bouman, J., Gerstl, M., et al. (2012). GGOS Bureau for standards and conventions: Integrated standards and conventions for geodesy. In S. C. Kenyon, M. C. Pacino, U. J. Marti (Eds.), *Geodesy for Planet Earth* (IAG Symposia, Vol. 136, pp. 995–998). doi:10.1007/978-3-642-20338-1_124.

Ihde, J., Wilmes, H., Müller, J., Denker, H., Voigt, C., & Hosse, M. (2010). *Validation of satellite gravity field models by regional terrestrial data sets, system Earth via geodetic-geophysical space techniques, Part 3* (pp. 277–296). Springer, Berlin: Advanced technologies in earth sciences. doi:10.1007/978-3-642-10228-8_22.

Intergovernmental Committee on Surveying and Mapping. (2002). *Geocentric datum of Australia technical manual, version 2.2*. http://www.anzlic.org.au/icsm/gdatm/

Jarecki, F., Wolf, I., Denker, H., & Müller, J. (2006). Quality assessment of GOCE gradients. In J. Flury, R. Rummel, C. Reigber, M. Rothacher, G. Boedecker, & U. Schreiber (Eds.), *Observation of the earth system from space* (pp. 271–286). Berlin: Springer.

Jiang, C. (Ed.). (1992). *Encyclopedia of China* (Solid-earth geophysics, surveying and mapping, space science). Beijing: Encyclopedia Of China Publishing House. 504pp (in Chinese). ISBN 7500059779.

Junkins, D. R. (1998). *Transforming to NAD83*. Report of the Geodetic Survey Division, Canada Centre for Surveying, Ottawa.

Karney, C. F. F. (2011). Transverse Mercator with an accuracy of a few nanometers. *Journal of Geodesy, 85*(8), 475–485.

Kawase, K. (2011). A general formula for calculating meridian arc length and its application to coordinate conversion in the Gauss-Krüger projection. *Bulletin of the Geospatial Information Authority of Japan, 59*, 1–13.

Kawase, K. (2012). Concise derivation of extensive coordinate conversion formulae in the Gauss-Krüger projection. *Bulletin of the Geospatial Information Authority of Japan, 60*, 1–6.

Kenyeres, A., Sacher, M., Ihde, J., Denker, H., & Marti, U. (2010). EUVN_DA: Realization of the European continental GPS/leveling network. In S. P. Mertikas (Ed.), *Gravity, geoid and earth observation* (IAG Symposia, Vol. 135, pp. 315–320). Berlin: Springer.

Kenyon, S. C., Pacino, M. C., & Marti, U. (2012). *Geodesy for planet earth* (IAG Symposia, Vol. 136). Berlin: Springer.

Klees, R., Revtova, E., Gunter, B., Ditmar, P., Oudman, E., Winsemius, H., et al. (2008). Filter design for GRACE gravity models. *Geophysical Journal International, 175*(2), 417–432.

Kong, X., & Mei, S. (2002). *Control survey* (2nd ed., Vol. 1). Wuhan: Wuhan University Press (in Chinese).

Krakiwsky, E. J., & Thomson, D. B. (1974, Reprinting 1995) *Geodetic position computations* (Technical Report, No. 217). Canada: University of New Brunswick, 99pp.

Krügel, M., & Angermann, D. (2007). Frontiers in the combination of space geodetic techniques. In P. Tregoning & C. Rizos (Eds.), *Dynamic planet* (IAG Symposia, Vol. 130, pp. 158–165). Heidelberg: Springer. doi:10.1007/978-3-540-49350-1_25.

Krügel, M., Angermann, D., Drewes, H., Gerstl, M., Meisel, B., & Tesmer, V. (2007a). *GGOS-D reference frame computations* (Geotechnologien Science Report No. 11). Potsdam: Koordinierungsbüro Geotechnologien. ISSN 1619-7399.

Krügel, M, Thaller, D., Tesmer, V., Rothacher, M., Angermann, D., & Schmid, R. (2007b). Tropospheric parameters: combination studies based on homogeneous VLBI and GPS data.

In H. Schuh, A. Nothnagel, C. Ma (Eds.), VLBI special issue. *Journal of Geodesy*, *81*(6/8), 515–527. doi:10.1007/s00190-006-0127-8

Krüger, L. (1912). *Konforme Abbildung des Erdellipsoids in der Ebene*. Leipzig: B.G. Teubner. 172pp.

Langley, R. B. (1992). Basic geodesy for GPS. *GPS World, 3*, 44–49.

Lee, L. P. (1946). The transverse mercator projection of the spheroid. *Survey Review, 8*(58), 142–152.

Legrand, J., Bergeot, N., Bruyninx, C., Wöppelmann, G., Bouin, M.-N., & Altamimi, Z. (2009). Impact of regional reference frame definition on geodynamic interpretations. *Journal of Geodynamics, 49*, 116–122. doi:10.1016/j.jog.2009.10.002.

Leick, A. (2004). *GPS satellite surveying* (3rd ed.). New York: Wiley.

Leuliette, E., Nerem, R., & Mitchum, G. (2004). Calibration of TOPEX/Poseidon and Jason altimeter data to construct a continuous record of mean sea level change. *Marine Geodesy, 27*, 79–94.

Li, J., Chen, J., Ning, J., & Chao, D. (2003). *Approximation theory of the earth gravity field and determination of China quasi-geoid 2000*. Wuhan: Wuhan University Press. 293pp (in Chinese). ISBN 9787307037144.

Li, G., Yang, Q., & Hu, D. (1993). *Map projection*. Beijing: Chinese People's Liberation Army Publishing House (in Chinese).

Liang, Z. (1999). Considerations of establishing China's unified height and sounding datums on land and sea. *Science of Surveying and Mapping, 3*, 20–29 (in Chinese).

Liu, J., Chen, J., Zhang, Y., Li, Y., Ge, M. (1999). *Wide area differential GPS principles and methods*. Beijing: Surveying and Mapping Press. 148pp (in Chinese). ISBN 7503009152.

Liu, J., & She, B. (1990). The transformation models for combining different types of 3-D positioning network. *Journal of Wuhan Technical University of Surveying and Mapping, 15* (2), 48–57 (in Chinese).

Liu, D., Shi, Y., & Guo, J. (1996). *Principles and data processing of GPS*. Shanghai: Tongji University Press (in Chinese).

Lu, Z. (1996). *Theories and methods of the Earth's gravity field*. Beijing: Chinese People's Liberation Army Press (in Chinese).

Lu, Z., Li, J., & Zhang, X. (2012). *China geodetic coordinate system 2000 and its application*. Beijing: PLA Press (in Chinese).

Lu, Z., Xu, D., & Zhu, H. (1995). The divide minimum method in geocentric coordinate transformation. *Acta Geodaetica et Cartographica Sinica, 24*(4), 275–282 (in Chinese).

Lu, Z., & Zhu, H. (1993). The testing and unified expression of coordinate transformation models. *Acta Geodaetica et Cartographica Sinica, 22*(3), 161–168 (in Chinese).

Maling, D. H. (1992). *Coordinate systems and map projections*. Oxford: Pergamon Press. 255pp.

Mertikas, S. P. (Ed.). (2010). *Gravity, geoid and Earth observation* (IAG Symposia, Vol. 135). Berlin: Springer.

Moore, A. W., & Neilan, R. E. (2005). The International GPS service tracking network: enabling diverse studies and projects through International cooperation. *Journal of Geodynamics, 40* (4/5), 461–469.

Morgan, J. G. (1987). The north American datum of 1983. *The Leading Edge, 6*(1), 18–25.

Moritz, H. (2000). The geodetic reference system 1980. *Journal of Geodesy, 74*(1), 128–133.

Moritz, H., & Mueller, L. I. (1987). *Earth rotation—Theory and observation*. New York: Ungar Publishers.

Müller, H., & Angermann, D. (2009). The international terrestrial reference frame—Latest developments. In *Proceedings of the 16th international workshop on laser ranging, Poznan, Poland* (Vol. 1, pp. 27–34).

Neilan, R. E., Moore, A., Dow, J., Gendt, G., & Weber, R. (2004). International GPS service—10 years history, new directions for GNSS and space geodesy. American Geophysical Union, Fall Meeting.

Neumeyer, J. (2010). Superconducting gravimetry. In G. Xu (Ed.), *Sciences of geodesy—I*. Berlin: Springer.

NGA. (2002). *Implementation of the World Geodetic System 1984 (WGS84) Reference Frame G1150. US National Geospatial-Intelligence Agency*, Addendum to NIMA TR 8350.2.

NIMA. (1989). *The Universal Grids: Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS)* (1st ed.). TM 8358.2, US National Imagery and Mapping Agency, Washington, DC.

NIMA. (1990). *Datums, ellipsoids, grids, and grid reference systems* (1st ed.). TM 8358.1, US National Imagery and Mapping Agency, Washington, DC.

NIMA. (2000). *Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems* (3rd ed.). US National Imagery and Mapping Agency Technical Report 8350.2.

Nin, J., Liu, J., Chen, J., et al. (2006). *Theories and technologies of modern geodesy*. Wuhan: Wuhan University Press (in Chinese).

Ning, J. (2002). Modern geodetic reference system. *Acta Geodaetica et Cartographica Sinica, 31*, 7–11 (in Chinese).

Ning, J., Chen, J., Li, d., Liu, J., & Zhang, Z. (2004). *Introduction of surveying and mapping*. Wuhan: Wuhan University Press. 283pp (in Chinese). ISBN 7307043807.

Office of the Surveyor-General. (2007). Standard for New Zealand geodetic datum 2000. Land Information New Zealand.

Pail, R., Bruinsma, S., Migliaccio, F., Förste, C., Goiginger, H., Schuh, W., et al. (2011). First GOCE gravity field models derived by three different approaches. *Journal of Geodesy, 85*, 819–843.

Paragi, Z., Reynolds, C., Biggs, A. D., Imai, H., & Garrett, M. A. (2005). The European VLBI network: Operations, reliability and performance, future directions in high resolution astronomy. In J. Romney, M. Reid (Eds.), *The 10th Anniversary of the VLBA, ASP Conference Proceedings* (Vol. 340, p. 611). San Francisco: Astronomical Society of the Pacific.

Pearlman, M. R., Degnan, J. J., & Bosworth, J. M. (2002). The international laser ranging service. *Advances in Space Research, 30*, 135–143.

Pent, G., & Luzum, B. (Eds.). (2010). *IERS conventions 2010* (IERS Technical Note No. 36). Frankfurt: Verlag des Bundesamts für Kartographie und Geodäsie. ISBN 3-89888-989-6, 179pp.

Petit G. (2000) Importance of a common framework for realization of space-time reference systems. In R. Rummel, H. Drewes, & W. Bosch, H. Hornik (Eds.) Towards an integrated global geodetic observing system (IAG Symposium, Vol. 120, pp. 3–7). IAG Symposia, Munich, October 1998.

Petrov, L., Gordon, D., Gipson, J., MacMillan, D., Ma, C., Fomalont, E., et al. (2009). Precise geodesy with the very long baseline array. *Journal of Geodesy, 83*(9), 859–876.

Plag, H. P., & Pearlman, M. (Eds.). (2009). *Global geodetic observing system—Meeting the requirements of a global society on a changing planet in 2020*. Berlin: Springer. ISBN 978-3-642-02686-7.

Ray, J., Dong, D., & Altamimi, Z. (2004). IGS reference frames: status and future improvements. *GPS Solutions, 8*, 4. doi:10.1007/s10291-004-0110-x.

Redfearn, J. C. B. (1948). Transverse Mercator formulae. *Survey Review, 9*(69), 318–322.

Reinking, J. (2010). Marine geodesy. In G. Xu (Ed.), *Sciences of geodesy—I* (pp. 275–299). Heidelberg: Springer. ISBN 3642117406.

Reinking, J., Smit-Philipp, H., & Even-Tzur, G. (2011). Surface deformation along the Carmel Fault System, Israel. *Journal of Geodynamics, 52*, 321–331.

Reiterer, A., Egly, U., Henert, M., & Riedel, B. (Eds.) (2010). *Application of artificial intelligence and innovations in engineering geodesy*. Second international workshop (AIEG 2010), Braunschweig, Germany, ISBN 978-3-9501492-6-5.

Rothacher, M., Angermann, D., Artz, T., Bosch, W., Drewes, H., Boeckmann, S., et al. (2011). GGOS-D: Homogeneous reprocessing and rigorous combination of space geodetic observations. *Journal of Geodesy, 85*(10), 679–705. doi:10.1007/s00190-011-0475-x.

Rothacher, M., Boonkamp, H., Angermann, D., & Böhm, J. (2007). *Sub-Commission 1.1. Coordination of space techniques, IAG Commission 1—reference frames*. Final report (IAG Commission 1 Bulletin No. 20, pp. 13–24). München: DGFI.

Sánchez, L., Seemüller, W., & Seitz, M. (2012). Combination of the weekly solutions delivered by the SIRGAS processing centre for the SIRGAS-CON reference frame. In S. Kenyon, M. C. Pacino, U. Marti (Eds.), *Geodesy for Planet Earth* (IAG Symposia, Vol. 136, pp. 845–851). doi:10.1007/978-3-642-20338-1_106.

Sánchez, L., & Seitz, M. (2011). *Recent activities of the IGS Regional Network Associate Analysis Centre for SIRGAS (IGS RNAAC SIR)* (DGFI Report No. 87).

SBSM. (2004). *Specifications for China National Triangulation (GBT 17942-2000)*. Beijing: State Bureau of Surveying and Mapping of China, The Standards Press of China (in Chinese).

SBSM. (2009). *Specifications for GPS Surveys (GBT 19314-2009)*. Beijing: State Bureau of Surveying and Mapping of China, The Standards Press of China (in Chinese).

Schluter, W., & Behrend, D. (2007). The international VLBI service for geodesy and astrometry (IVS): Current capabilities and future prospects. *Journal of Geodesy, 81*(6/8), 379–387.

Schreiber, O. (1866). *Theorie der Projektionsmethode der hannoverschen Landsvermessung*. Hannover: Hahnsche Hofbuchhandlung.

Schwarz, C. (Ed.). (1989). *North American datum of 1983*. NOAA Professional Paper NOS 2. Rockville, MD: National Geodetic Information Center, NOAA, 256pp.

Schwarz, K. P. (Ed.). (2000). *Geodesy beyond 2000—The challenges of the first decade* (IAG Symposia, Vol. 121). Berlin: Springer. ISBN 3540670025.

Seeber, G. (2003). *Satellite geodesy* (2nd ed.). Berlin: Walter de Gruyter. 589pp. ISBN 3-11-017549-5.

Seemüller, W., Seitz, M., Sánchez, L., & Drewes, H. (2012). The new multi-year position and velocity solution SIR09P01 of the IGS Regional Network Associate Analysis Centre (IGS RNAAC SIR). In S. Kenyon, M. C. Pacino, U. Marti (Eds.), *Geodesy for planet Earth* (IAG Symposia, Vol. 136, pp. 877–883). doi:10.1007/978-3-642-20338-1_110.

Seitz, M. (2012a). Comparison of different combination strategies applied for the computation of terrestrial reference frames and geodetic parameter series. In *Proceedings of the 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS) 2011*, Munich.

Seitz, M. (2012b). *Report on the IAG Symposia G01: Reference frames from regional to global scales*. IUGG Symposia 2012, Melbourne. zfv, 137/1.

Seitz, M., Angermann, D., Bloßfeld, M., Drewes, H., & Gerstl, M. (2012a). The 2008 DGFI realization of the ITRS: DTRF2008. *Journal of Geodesy, 86*(12), 1097–1123. doi:10.1007/s00190-012-0567-2.

Seitz, M., Angermann, D., Bloßfeld, M., Gerstl, M., Heinkelmann, R., & Müller, H. (2012b). ITRS Combination Centre at DGFI. In *IERS Annual Report 2011*, Frankfurt a.M.: BKG.

Seitz, M., Angermann, D., & Drewes, H. (2013). Accuracy assessment of ITRS 2008 realization of DGFI: DTRF2008. In Z. Altamimi & X. Collilieux (Eds.), *Reference frames for applications in geosciences* (International Association of Geodesy Symposia, Vol. 138, pp. 87–93). Berlin: Springer. doi:10.1007/978-3-642-32998-2_15.

Seitz, M., Angermann, D., Drewes, H., Gerstl, M., Heinkelmann, R., & Müller, H. (2011). ITRS Combination Centre at DGFI. In *IERS Annual Report 2008/2009* (pp. 127–132). Frankfurt a. M.: BKG.

Sella, G. F., Dixon, T. H., & Mao, A. (2002). REVEL: A model for recent plate velocities from space geodesy. *Journal of Geophysical Research, 107*(B4), 2081. doi:10.1029/2000JB00033. 1–30.

Sideris, M. G. (Ed.). (2009). *Observing our changing Earth* (IAG Symposia, Vol. 133). Berlin: Springer.

Sillard, P., & Boucher, C. (2001). A review of algebraic constraints in terrestrial reference frame definition. *Journal of Geodesy, 75*(2/3), 63–73.

Sjöberg, L. E. (2006). New solutions to the direct and indirect geodetic problems on the ellipsoid. *Zeitschrift fuer Vermessungswesen, 131*, 35–39.

Sjöberg, L. E., & Shirazian, M. (2012). Solving the direct and inverse geodetic problems on the ellipsoid by numerical integration. *Journal of Surveying Engineering, ASCE, 138*(1), 9–16.

Steed, J. (1990). A practical approach to transformation between commonly used reference systems. *The Australia Surveyor, 35*(3), 248–264.

Takahashi, F., Kondo, T., Takahashi, Y., & Koyama, Y. (2000). *Very long baseline interferometer* (Wave Summit Course). Amsterdam: Ohmsha, IOS Press.

Thaller, D., Dach, R., Seitz, M., Beutler, G., Mareyen, M., & Richter, B. (2011). Combination of GNSS and SLR observations using satellite co-locations. *Journal of Geodesy, 85*(5), 257–272. doi:10.1007/s00190-010-0433-z.

Thaller, D., Krügel, M., Rothacher, M., Tesmer, V., Schmid, R., & Angermann, D. (2007). Combined Earth orientation parameters based on homogeneous and continuous VLBI and GPS data. In H. Schuh, A. Nothnagel, C. Ma (Eds.), VLBI special issue. *Journal of Geodesy*, *81*(6/8), 529–541.

Thomas, C. M., & Featherstone, W. E. (2005). Validation of Vincenty's formulas for the geodesic using a new fourth-order extension of Kivioja's formula. *Journal of Surveying Engineering, 131*(1), 20–26.

Thomson, D. B. (1976). *Combination of geodetic networks*. Canada: University of New Brunswick (Technical Report No. 30, p. 198).

Timmen, L. (2003). Precise definition of the effective measurement height of free-fall absolute gravimeters. *Metrologia, 40*, 62–65.

Timmen, L. (2010). Absolute and relative gravimetry. In G. Xu (Ed.), *Sciences of geodesy—I: Advances and future directions* (pp. 1–48). Berlin: Springer.

Timmen, L., Falk, R., Gitlein, O., & Wilmes, H. (2011). The measuring offset of the absolute gravimeter JILAg-3 (LUH) with respect to the FG5 instruments no. 101 (BKG) and no. 220 (LUH). In V. G. Peshekhonov (Ed.), *Terrestrial gravimetry: static and mobile measurements, international symposium* (pp. 72–77). Saint Petersburg: The State Research Center of the Russian Federation—Concern CSRI Elektropribor, JCS.

Timmen, L., Flury, J., Peters, T., & Gitlein, O. (2006). A new absolute gravity base in the German Alps. In M. Hvoždara & I. Kohún (Eds.), *Contributions to geophysics and geodesy, 2nd workshop on international gravity field research (special issue)* (Vol. 36, pp. 7–20). Bratislava: Slovak Academy of Sciences.

Timmen, L., Gitlein, O., Klemann, V., & Wolf, D. (2009). *Gravity change in the Fennoscandian Uplift Area observed by absolute gravimetry* (Vol. 11, p.1834). EGU General Assembly Conference Abstracts.

Timmen, L., Gitlein, O., Klemann, V., & Wolf, D. (2012). Observing gravity change in the Fennoscandian uplift area with the Hanover absolute gravimeter. *Pure and Applied Geophysics, 169*(8), 1331–1342.

Timmen, L., & Müller, J. (2010). Accuracy and time stability of the Hannover absolute gravimeters JILAg-3 and FG5-220. In *The apple of knowledge* (pp. 251–259). Thessaloniki: Ziti Editions.

Titov, O. (2010). VLBI2020: from reality to vision. In D. Behrend, K. D. Baver (Eds.), *IVS 2010 General meeting proceedings* (pp. 60–64). NASA/CP-2010-215864.

Tobita, M. (2001). Coordinate transformation software "TKY2JGD" from Tokyo datum to a geocentric reference system, Japanese Geodetic Datum 2000. *Times of Geographical Survey Institute, 97*, 31–57.

Torge, W. (1989). *Gravimetry*. Berlin: Walter de Gruyter.

Torge, W., & Müller, J. (2012). *Geodesy* (4th ed.). Berlin: Walter de Gruyter. 433pp. ISBN 978-3-11-020718-7.

Van Camp, M., Williams, S. D. P., & Francis, O. (2005). Uncertainty of absolute gravity measurements. *Journal of Geophysical Research, 110*, B05406. doi:10.1029/2004JB003497.

Vanicek, P., & Krakiwsky, E. J. (1986). *Geodesy: The concepts*. Amsterdam: North-Holland. 697pp. ISBN 978-0-444-86149-8.

Vermeille, H. (2002). Direct transformation from geocentric coordinates to geodetic coordinates. *Journal of Geodesy, 76*, 451–454.

Vermeille, H. (2004). Computing geodetic coordinates from geocentric coordinates. *Journal of Geodesy, 78*, 94–95.

Vincenty, T. (1976). Direct and inverse solutions of geodesics on the ellipsoid with applications of nested tables. *Survey Review, 23*(176), 88–93.

Voigt, C., Denker, H., & Hirt, C. (2009). Regional astrogeodetic validation of GPS and levelling data and quasigeoid models. In M. Sideris (Ed.), *Observing our changing earth, Proceedings of the IAG General Assembly*, Perugia, 2–13 July 2007 (IAG Symposia, Vol. 133, pp. 413–420).

Wastson, C., Tregoing, P., & Coleman, R. (2006). The impact of solid earth tide models on GPS time series analysis. *Geophysical Research Letters, 33*(8), L08306. doi:10.1029/2005GL025538.

Wei, Z. (2003). National geodetic coordinate system: to next generation. *Geomatics and Information Science of Wuhan University, 28*(2), 138–143 (in Chinese).

Wei, Z., & Ge, M. (1998). *Relative positioning mathematical model of GPS*. Beijing: Surveying and Mapping Press. 182pp (in Chinese). ISBN 9787503009099.

Wöppelmann, G., Bouin, M.-N., & Altamimi, Z. (2008). Terrestrial reference frame implementation in global GPS analysis at TIGA ULR consortium. *Physics and Chemistry of the Earth, 3–4* (33), 217–224. doi:10.1016/j.pce.2006.11.001.

Wöppelmann, G., Letetrel, C., Santamaría, Á., Bouin, M.-N., Collilieux, X., Altamimi, Z., et al. (2009). Rates of sea-level change over the past century in a geocentric reference frame. *Geophysical Research Letters, 36*, L12607. doi:10.1029/2009GL038720.

Wöppelmann, G., Martín Míguez, B., Bouin, M.-N., & Altamimi, Z. (2007). Geocentric sea-level trend estimates from GPS analyses at relevant tide gauges world-wide. *Global and Planetary Change, 57*(3–4), 396–406. doi:10.1016/j.gloplacha.2007.02.002.

Wu, X., Collilieux, X., Altamimi, Z., Vermeersen, B., Gross, R. S., & Fukumori, I. (2011). Accuracy of the international terrestrial reference frame origin and earth expansion. *Geophysical Research Letters, 38*, L13304. doi:10.1029/2011GL047450.

Wu, X., Ray, J., & van Dam, T. (2012). Geocenter motion and its geodetic and geophysical implication. *Journal of Geodesy, 58*, 44–61.

Xia, Y., & Huang, T. (1995). *Spherical astronomy*. Nanjing: Nanjing University Press (in Chinese).

Xiong, J. (1985). Ancient arc measurement in China. *Acta Geodaetica et Cartographica Sinica, 14* (4) (in Chinese).

Xiong, J. (1989). *Ellipsoidal geodesy*. Beijing: Chinese People's Liberation Army Publishing House (in Chinese).

Xu, Q. (2001). *Space geodesy*. Beijing: Chinese People's Liberation Army Publishing House (in Chinese).

Xu, G. (2003). *GPS—Theory, algorithms and application*. Heidelberg: Springer.

Xu, G. (2007). *GPS—Theory, algorithms and applications* (2nd ed.). Heidelberg: Springer. 350pp.

Xu, G. (Ed.). (2010). *Sciences of geodesy—I: Advances and future directions*. Berlin: Springer.

Xu, H. (2010b). *Solid earth tides*. Wuhan: Hubei Science and Technology Press. 304pp (in Chinese). ISBN 9787535245441.

Xu, G. (Ed.). (2012). *Sciences of geodesy—II: Innovations and future developments*. Berlin: Springer.

Xu, Z., Liu, Z., & Wu, G. (1991). *Geodetic control surveys*. Beijing: Chinese People's Liberation Army Publishing House (in Chinese).

Xu, G., Timmen, L., & Bastos, L. (1997). GPS kinematic positioning in AGMASCO campaigns-strategic goals and numerical results. In *Proceedings of the 10th international technical*

meeting of the Satellite Division of The Institute of Navigation (ION GPS 1997) (pp. 1173–1183).

Xu, S., Zhang, H., & Yang, Z. (1998). *Principles and applications of GPS surveying*. Wuhan: Wuhan Technical University of Surveying and Mapping Press (in Chinese).

Yang, Y. (2009). Chinese geodetic coordinate system 2000. *Chinese Science Bulletin, 2009*(54), 2714–2721.

Yang, Y., Tang, Y., Chen, C., Wang, M., Zhang, P., Wang, X., et al. (2007). National 2000 GPS control network of China. *Progress in Natural Science, 17*(8), 983–987.

Yang, Y., Tang, Y., Chen, C., Wang, M., Zhang, P., Wang, X., et al. (2009). Integrated adjustment of Chinese 2000 GPS control network. *Survey Review, 41*(313), 226–237.

Yang, Y., Zeng, A., & Wu, F. (2011). Horizontal crustal movement in China fitted by adaptive collocation with embedded euler vector. *Science China: Earth Sciences, 54*(12), 1822–1829.

Yang, Y., Zha, M., Song, L., Wei, Z., Wang, Z., Ouyang, G., et al. (2005). Combined adjustment project of national astronomical geodetic network and 2000 national GPS control network. *Progress in Natural Science, 15*(5), 435–441.

Ye, S., & Huang, C. (Eds.). (2000). *Astrogeodynamics*. Jinan: Shandong Science and Technology Press (in Chinese).

Zhang, Q., & Liang, Z. (2001). Model updating of China national vertical control network. *Bulletin of Surveying and Mapping, 3* (in Chinese).

Zhu, H. (1986). *The establishment of geodetic coordinate system*. Beijing: Surveying and Mapping Press (in Chinese).

Zhu, H., & Huang, J. (1993). *Geodetic computation on the ellipsoid*. Beijing: Chinese People's Liberation Army Publishing House (in Chinese).

Index

Z. Lu et al., *Geodesy*, DOI 10.1007/978-3-642-41245-5,