ASTRONOMY AND ASTROPHYSICS LIBRARY

Gerhard Beutler

Methods of Celestial Mechanics

Volume I: Physical, Mathematical, and Numerical Principles

In Cooperation with
Prof. Leos Mervart and Dr. Andreas Verdun

With 99 Figures Including 11 Color Figures,
32 Tables and a CD-ROM

Springer

Professor Dr. Gerhard Beutler

Universität Bern
Astronomisches Institut
Sidlerstrasse 5
3012 Bern, Switzerland
e-mail: gerhard.beutler@aiub.unibe.ch

Cover picture: This mosaic shows images of Earth and the moon acquired by the multispectral imager on the Near Earth Asteroid Rendezvous Spacecraft (NEAR) on January 23, 1998, 19 hours after the spacecraft swung by Earth on its way to the asteroid 433 Eros. The images of both were taken from a range of 250,000 miles (400,000 km), approximately the same as the distance between the two bodies. This unique perspective, never seen before, shows both our planet and its moon at the relative size that each appears when viewed from the other. Also, both Earth and the Moon are viewed from above their south poles, a perspective not attainable from either body because the moon orbits high above Earth's equator. In the Earth image, the south pole is at the center and the continent of Antarctica is surrounded by sea ice and storm fronts. The image mosaic is contructed from blue, green, and infrared filters. These colors highlight differences between rock types, water, and vegetation. On Earth, the red area at the upper right side is desert and vegetation in Australia. Snow, ice, and clouds appear as sublty different shades of white and light blue. The moon's blandness, compared to Earth, arises from its lack of an atmosphere, oceans, and vegetation. For viewing purposes, the moon is shown five times brighter than in reality, and ten times closer to Earth than it actually is.
Built and managed by The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, NEAR was the first spacecraft launched in NASA's Discovery Program of low-cost, small-scale planetary missions.

Library of Congress Control Number: 2004105868

ISSN 0941-7834
ISBN 3-540-40749-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Typesetting by the author
Final layout: Frank Herweg, Leutershausen
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper 55/3141/ba - 5 4 3 2 1 0

A&A LIBRARY

ASTRONOMY AND ASTROPHYSICS LIBRARY

The Stars By E. L. Schatzman and F. Praderie

Modern Astrometry 2nd Edition
By J. Kovalevsky

The Physics and Dynamics of Planetary Nebulae By G. A. Gurzadyan

Galaxies and Cosmology By F. Combes, P. Boissé, A. Mazure and A. Blanchard

Observational Astrophysics 2nd Edition
By P. Léna, F. Lebrun and F. Mignard

Physics of Planetary Rings Celestial Mechanics of Continuous Media
By A. M. Fridman and N. N. Gorkavyi

Tools of Radio Astronomy 4th Edition
By K. Rohlfs and T. L. Wilson

Astrophysical Formulae 3rd Edition (2 volumes)
Volume I: Radiation, Gas Processes and High Energy Astrophysics
Volume II: Space, Time, Matter and Cosmology
By K. R. Lang

Tools of Radio Astronomy Problems and Solutions By T. L. Wilson and S. Hüttemeister

Galaxy Formation By M. S. Longair

Astrophysical Concepts 2nd Edition
By M. Harwit

Astrometry of Fundamental Catalogues The Evolution from Optical to Radio Reference Frames
By H. G. Walter and O. J. Sovers

Compact Stars. Nuclear Physics, Particle Physics and General Relativity 2nd Edition
By N. K. Glendenning

The Sun from Space By K. R. Lang

Stellar Physics (2 volumes)
Volume 1: Fundamental Concepts and Stellar Equilibrium
By G. S. Bisnovatyi-Kogan

Stellar Physics (2 volumes)
Volume 2: Stellar Evolution and Stability
By G. S. Bisnovatyi-Kogan

Theory of Orbits (2 volumes)
Volume 1: Integrable Systems and Non-perturbative Methods
Volume 2: Perturbative and Geometrical Methods
By D. Boccaletti and G. Pucacco

Black Hole Gravitohydromagnetics
By B. Punsly

Stellar Structure and Evolution
By R. Kippenhahn and A. Weigert

Gravitational Lenses By P. Schneider, J. Ehlers and E. E. Falco

Reflecting Telescope Optics (2 volumes)
Volume I: Basic Design Theory and its Historical Development. 2nd Edition
Volume II: Manufacture, Testing, Alignment, Modern Techniques
By R. N. Wilson

Interplanetary Dust
By E. Grün, B. Å. S. Gustafson, S. Dermott and H. Fechtig (Eds.)

The Universe in Gamma Rays
By V. Schönfelder

Astrophysics. A New Approach 2nd Edition
By W. Kundt

Cosmic Ray Astrophysics
By R. Schlickeiser

Astrophysics of the Diffuse Universe
By M. A. Dopita and R. S. Sutherland

The Sun An Introduction. 2nd Edition
By M. Stix

Order and Chaos in Dynamical Astronomy
By G. J. Contopoulos

Astronomical Image and Data Analysis
By J.-L. Starck and F. Murtagh

ASTRONOMY AND ASTROPHYSICS LIBRARY

The Early Universe Facts and Fiction
4th Edition By G. Börner

The Design and Construction of Large Optical Telescopes By P. Y. Bely

The Solar System 4th Edition
By T. Encrenaz, J.-P. Bibring, M. Blanc, M. A. Barucci, F. Roques, Ph. Zarka

General Relativity, Astrophysics, and Cosmology By A. K. Raychaudhuri, S. Banerji, and A. Banerjee

Stellar Interiors Physical Principles, Structure, and Evolution 2nd Edition
By C. J. Hansen, S. D. Kawaler, and V. Trimble

Asymptotic Giant Branch Stars
By H. J. Habing and H. Olofsson

The Interstellar Medium
By J. Lequeux

Methods of Celestial Mechanics (2 volumes)
Volume I: Physical, Mathematical, and Numerical Principles
Volume II: Application to Planetary System, Geodynamics and Satellite Geodesy
By G. Beutler

To my family
and my friends and co-workers
at the Astronomical Institute of the University of Bern

Preface

This work is based on lecture notes about Celestial Mechanics of the planetary system and of artificial satellites, about the rotation of Earth and Moon, and about numerical analysis. The lectures were intended for diploma students of astronomy, physics, mathematics, and geography at the University of Bern in their first three academic years. In view of the broad and inhomogeneous audience, the lectures had to be self-consistent and based on simple, generally known physical and mathematical facts and concepts.

The work consists of three parts in two volumes, where the first Volume is reserved for the development of theory and mathematical tools (Part I), the second for applications (Part II) and the accompanying computer program system (Part III). The two volumes were designed and written as one self-contained work. Logically, they belong together. The program system, Part III, is used throughout the two first parts to illustrate the theoretical concepts in Volume I and the applications in Volume II.

The two volumes offer a thorough introduction into modern astrodynamics and its applications to a broad variety of problems in the planetary system and in global geodynamics for students in astronomy, physics, geodesy (in particular space geodesy), geophysics and geosciences, and in applied mathematics. They contain many ideas for future research in the areas addressed, and they are hopefully also beneficial for the experts in the field.

The equations of motion – for point masses and extended bodies – and their solution methods are the central, unifying aspect of our treatment of Celestial Mechanics and its applications. Particular solutions of the equations of motion, specified, e.g., by initial conditions and by the parameters defining the force field acting on the celestial bodies considered, are studied with the tools of numerical analysis. Powerful and yet efficient numerical algorithms are developed, starting from basic mathematical principles. The focus of our treatment of numerical analysis is not on completeness but on a few key methods, which are based on simple and easily understandable mathematical principles.

We tried to avoid concepts which are “only” useful to the specialist in the field, be it in numerical analysis or in Celestial Mechanics. If the choice was either that of an elegant or of a simple treatment, the simple approach was followed.

Both volumes are accompanied by the same Compact Disk (CD) containing the computer programs as executable modules for Personal Computers (PC). The program system is easy to install and to use on PCs with a Windows operating system. The computer programs are documented in Part III, Volume II. The programs are a central issue of our treatment of astrodynamics, not just as a nice supplement. They allow it to study and analyze a wide variety of problems reaching far beyond those treated in this book. The program package should be useful for academic teachers and their students, but also for research workers.

Prof. Leoš Mervart of the Technical University of Prague designed and wrote the menu system accompanying the computer programs. It is in essence his merit that the computer-programs are easy to understand and to use. Part III of Volume II were written jointly by him and the undersigned. In addition, Leoš Mervart was proof-reading Chapter 3 (Equations of Motion) of the main text and he was producing most of the more delicate figures.

My colleague and co-worker Dr. Andreas Verdun was the design expert concerning the structure and the formal appearance of this work. In addition, his collaboration was paramount in all aspects related to his specialization, the history of astronomy, in particular of Celestial Mechanics. He screened and proof-read the entire manuscript. His expertise and never ending encouragement was of greatest importance for the realization and completion of this work.

This work never could have been completed without the assistance of the two young colleagues. Their contribution is acknowledged with deep gratitude.

Prof. Paul Wild, my predecessor as director of the Astronomical Institute of the University of Bern (AIUB), contributed in many respects to this book. The observations of minor planets used as examples in the chapter about orbit determination were performed and reduced by Paul Wild personally. The observations of minor planets and comets provided on the CD mostly refer to objects discovered by him at the Zimmerwald Observatory. Paul Wild also adapted his fabulous skill to screen Schmidt-plates for new objects (minor planets, comets, supernovae, etc.) to the manuscript of this book by performing an amazingly thorough proof-reading of major parts of the manuscript. The final result is undoubtedly very much improved thanks to his effort.

Chapters 4 (two- and three-body problems) and 7 (numerical analysis) were proof-read by Adrian Jäggi, Chapter II- 2 (rotation of Earth and Moon) by Claudia Urschl, and Chapter II- 3 (satellite motion) by Michael Meindl. The three young colleagues are Ph.D.-candidates at our institute. Chapter 8 (orbit determination and parameter estimation) was proof-read by Dr. Thomas Schildknecht, head of the institute's CCD astrometry group. He also reviewed the Chapter II- 4 (planetary system). Dr. Urs Hugentobler received his diploma in theoretical physics, then joined the CCD group and wrote a Ph.D. thesis in the field of astrometry and Celestial Mechanics. After a longer research stay at ESOC in Darmstadt, he joined the AIUB team as head of AIUB GPS research group. With his broad background and his sharp mind he was perfectly suited to proof-read the entire Part II and the chapter about orbit determination of Volume I.

Profs. Robert Weber from the Technical University of Vienna, Markus Rothacher from the Technical University of Munich, and Prof. Werner Gurtner, director of the Zimmerwald Observatory, also read and commented parts of the manuscript. Dr. Jan Kouba from the National Geodetic Survey of Canada thoroughly read most of Part II. The comments by the four distinguished colleagues are very much appreciated. A final proof-reading at the entire manuscript was performed by Ms Edith Stöveken and Ms Claudia Urschl.

The editing and reviewing process of a treatise of this extent is a crucial aspect, at times even a nightmare. The reviewing work was a considerable addition to the normal professional duties of the colleagues mentioned above and to those of the author. It is my sincere desire to thank my friends and colleagues for their assistance. I can only promise to assist them in a similar way, should they decide to achieve something similar. I cannot recommend this to anybody, on the other hand: My sabbatical leave from the University of Bern in spring and summer 2001 and the following two years were in essence sacrificed to the purpose of writing and completing this two volume work.

The author hopes that the two volumes will be helpful to and stimulating for students and researchers – which in turn would help him to forget the "(blood), sweat and tears" accompanying the creative act.

Bern, February 2004 *Gerhard Beutler*

Contents

Part I. Physical, Mathematical, and Numerical Principles

Contents of Volume II

Part II. Applications

Part III. Program System

Part I

Physical, Mathematical, and Numerical Principles

1. Overview of the Work

This work contains three parts in two volumes. Volume I consists of Part I *Theory*, Volume II of Part II *Applications* and Part III *Program System.* Part I *Theory* contains Chapters 3 to 8, Part II *Applications* the Chapters II- 2 to II- 4, and Part III *Program System*, the documentation of the entire program system.

Celestial Mechanics, as we understand it today, has a long history. Its impact on the concepts of physics and mathematical analysis and, more recently, on geophysics can hardly be overestimated. Chapter 2 reviews the development of classical Celestial Mechanics, but also the developments related to the motion of artificial satellites.

1.1 Part I: Theory

The Equations of Motion. The title of this work demands that we depart from the equations of motion for the celestial bodies, which may be considered either as point masses or as extended bodies in our developments. In the latter case the bodies may be either rigid or deformable. This initial problem description is suited to address a great variety of problems: the orbital motion in galaxies, globular clusters, planetary systems, binaries, the orbital and rotational motion of planets, and the motion of natural and artificial satellite systems around a planet. Such a general treatment of Celestial Mechanics would be demanding, it could, however, hardly be dealt with in only two volumes.

We focus our treatise on the planetary system (consisting of a limited number of about $N \leq 20$ of point masses), on the orbital and rotational motion of the Earth-Moon-Sun system as an example of an N-body problem with extended bodies, and on the orbital motion of artificial Earth satellites (the attitude of satellites is briefly addressed, as well). With this selection of topics we leave aside many fascinating and important problems in dynamical astronomy, in particular the entire field of galactic dynamics. The latter topic is, e.g., very nicely treated in the standard textbook by Binney and Tremaine [20]. Our selection of topics still is rather ambitious, however.

The equations of motion for three types of problems, namely for the planetary system, for the three-body problem Earth-Moon-Sun and for the motion of artificial satellites are derived in Chapter 3. The method, simple and transparent in principle, is the same for all problem types, for the orbital as well as for the rotational motion: Using classical mechanics, Newton's law of universal gravitation, and the Newton-Euler formalism equating the time-derivatives of the linear momentum of individual bodies (or of mass elements of extended bodies) with the forces acting on the particles, the equations for the orbital and rotational motion are obtained in the inertial system. Depending on the problem type, the equations are then transformed to refer to the primary body for the particular problem type. When the motions of planets, minor planets or comets are studied, the position vectors in the equations of motion are heliocentric, in the other two cases, these position vectors are geocentric.

The three problem types have certain peculiarities: (1) "only" the orbital motion needs to be addressed in the first case, orbital *and* rotational motion in the second case; (2) "only" gravitational forces must be considered in the first two cases, whereas non-gravitational forces have to be dealt with, as well, in the third case; (3) due to the artificial satellites' proximity to the Earth, the gravitational potential of the Earth needs to be modelled very accurately; (4) in the latter application the equations of motion of different satellites are not coupled mathematically, allowing it to deal with each satellite orbit separately.

The developments are in essence based on classical mechanics. The relativistic equations of motion are, however, reproduced, as well. In Chapter 3 the so-called PPN (Parametrized Post-Newtonian) version of the equations of motion for the N-body problem is introduced and discussed, as well. The PPN equations may be viewed as a perturbation of the classical N-body motion. The direct use of the PPN equations (e.g., used for the production of planetary and lunar ephemerides in [107]) in numerical integrations over millions of years prohibitively affects the efficiency of the solution algorithms. This is why approximations of the correct PPN equations are considered, as well (and implemented as options into the computer programs for the planetary and satellite motion).

The Classical Two- and Three-body Problems. The two-body problem must be an integral part of each treatise of Celestial Mechanics, and it is found as an important issue in many textbooks of "ordinary" mechanics. Here, it is dealt with in Chapter 4 together with the three-body problem. That chapter thus deals with the two presumably simplest problems encountered in Celestial Mechanics.

The two-body problem, the motion of two point masses w.r.t. each other, may be solved "analytically", i.e., in terms of a finite set of elementary mathematical functions of time. These analytical solutions are extremely important as *first approximations* of more complicated problems. In Celestial Mechanics the two-body motion often is referred to as the *unperturbed* motion, implying that all other motions may be viewed as perturbed two-body motions. The chapter introduces the one-to-one relationship between the set of position- and velocity-vectors on one hand, and the orbital elements on the other hand. This relationship allows to introduce the concept of *osculating orbital elements*, by assigning one set of six orbital elements to each set of position- and velocity-vectors of a perturbed motion, using the formulae of the two-body problem.

Osculating and mean elements – the latter defined as averages of osculating elements over certain time intervals – are introduced as fundamental concepts. The computation of ephemerides, one of the important practical problems in Celestial Mechanics, is briefly addressed, as well. The three-body problem already contains many (if not most) of the characteristics and difficulties of the general planetary N-body motion. It was studied by many eminent astronomers and mathematicians (from Euler to Poincaré). The attempts to find "analytical" solutions of the 3-body problem were only moderately successful. They led, however, to the discovery of the problème restreint, one of the most charming mathematical miniatures found in dynamical astronomy. It is treated as a preparation to more general problems.

Variational Equations. The trajectory of a celestial body contains much information. Studies of the development of the osculating orbital elements as a function of time are indeed extremely informative, but yet it is impossible to decide in an objective way whether or not the findings are representative for other trajectories with similar initial characteristics. In order to answer such questions it is mandatory to study the so-called *variational equations*, which may be associated with each individual trajectory. The variational equations are of greatest importance in Celestial Mechanics – in theory as well as in application. They are required for orbit determination and for solving more general parameter estimation problems, in questions concerning the stability of a particular solution, and in error propagation studies. Chapter 5 introduces the variational equations as linear differential equations for the partial derivatives of the position vector(s) of celestial bodies w.r.t. the parameters defining the particular solution of the equations of motion considered. The chapter also provides analytical solutions (in the sense mentioned above) of the variational equations associated with the equations of motion of the two-body problem, and compares their characteristics with the solution characteristics related to perturbed motion.

Theory of Perturbations. Perturbation theory is the central topic of Chapter 6. Each method to solve an initial value problem associated with the equations of motion of a particular orbital motion may be viewed as a *perturbation method.* Usually one expects, however, that perturbation methods make (intelligent) use of the known approximative solutions, i.e., of the solutions of the corresponding two-body motion. The knowledge of an approximative solution may be exploited in many different ways. It is, e.g., possible to set up a differential equation for the difference vector between the actual solution and the known two-body approximation. The best-known of these attempts is the so-called *Encke method*, which is analyzed and considerably expanded when introducing the problems in the chapter. The best possible way to exploit the known analytical solution of the two-body problem consists of the derivation of differential equations for the osculating orbital elements. These equations are derived in an elementary way, without making use of the results of analytical mechanics. Our approach first leads to the perturbation equations in the Gaussian form, which allow it to consider a very broad class of perturbing functions. Only afterwards we derive the so-called Lagrange planetary equations, requiring the perturbing functions to be gradients of a scalar (so-called perturbation) function. The method to derive the equations is very simple and transparent, and the general form of the equations is amazingly simple. The drawback lies in the necessity to calculate the gradients of the orbital elements (w.r.t. Cartesian position- and velocity-components), a task which was performed in the last, technical section of the chapter.

When comparing the mathematical structure of the perturbation equations for different orbital elements (either in the Gaussian or in the Lagrangian form), one finds that all except one are essentially of the same simple mathematical structure. The exception is the equation for the time of pericenter passage T_0 (alternatively for the mean anomaly σ_0 referred to the initial epoch t_0), because the time argument figures outside the trigonometric functions on the right-hand sides of the equations. This is a nuisance independently of whether one solves the equations analytically or numerically. As opposed to the usual method of introducing new, auxiliary functions (as, e.g., the function ρ introduced by Brouwer and Clemence [27]), we derive directly a differential equation for the mean anomaly $\sigma(t)$ which does not show the problems mentioned above. $\sigma(t)$ is of course not an orbital element (an integration constant of the two-body motion), but any other auxiliary functions that might be introduced are not first integrals either.

Numerical Solution of Ordinary Differential Equations. Numerical analysis, in particular the numerical solution of the equations of motion and the associated variational equations, is studied in considerable detail in Chapter 7. In view of the fact that first- and second-order equations as well as definite integrals have to be solved in Celestial Mechanics, the general problem of solving non-linear differential equation systems of order n is studied first. Linear systems and integrals may then be considered as special cases of the general problem.

It is not sufficient to consider only initial value problems in Celestial Mechanics. The so-called *local boundary value problem* (where the boundary epochs are close together in time) is of particular interest. It is, e.g., used in orbit determination problems. Euler's original analysis (see Figure 7.1) is the foundation for all modern algorithms. It meets all requirements an algorithm should offer, except one: Euler's method is prohibitively inefficient. We show that the (not so well known) collocation methods may be viewed as the logical generalization of Euler's method. *Cum grano salis* one might say that the Euler method and the collocation method are identical except for the order of the approximation. Euler's method corresponds to a local Taylor series approximation of order n (where n is the order of the differential equation system); the order $q \geq n$ of the collocation method may be defined by the program user. Collocation methods may be easily adapted to automatically control the local errors of the integration, allowing it to determine efficiently not only orbits of small, but also large eccentricities.

Local error control is one issue, the accumulation of the local errors, due to two different sources, is another one. The accumulation of errors is of course studied for a machine-environment (as opposed to hand calculation). Apart from that our treatment is closely related to the method described in Brouwer's brilliant analysis [26]. Based on this theory a rule of thumb is provided for selecting the proper (constant) stepsize for producing planetary ephemerides (assuming low eccentricity orbits). This approximate treatment of the accumulation error is not applicable to very long integration spans or to problems involving strong perturbations (as, e.g., in the case of resonances). The correct theory of error accumulation therefore must to be based on the variational equations as well.

Orbit Determination and Parameter Estimation. Orbit determination and more general parameter estimation procedures are the topic of Chapter 8. The decision to conclude Part I with the chapter on orbit determination is justified by the fact that the problem reveals many interesting theoretical aspects related to parameter estimation theory. The determination of orbits may, however, also be viewed as one of the important practical tasks in Celestial Mechanics. The chapter may thus also be viewed as a transition chapter to the application part.

Orbits of celestial bodies may only be determined if they were repeatedly observed. For generations of astronomers the expression "observation" was synonymous for "direction observation" (usually an astrometric position), defining (in essence) the unit vector from the observer to the observed object at the epoch of the observation (a precise definition is provided). Except for the fact that today usually CCD (Charge Coupled Device) observations, and no longer photographic or even visual observations, are made, not much has changed in this view of things, when minor planets or comets are concerned. Orbit determination based on astrometric positions is also an important issue when dealing with artificial satellites and/or space debris. This is why the classical orbit determination problem based on astrometric positions applied to minor planets, comets, and artificial objects orbiting the Earth is addressed first.

It is important to distinguish between *first orbit determination* and *orbit improvement*. In the former case there is no a priori information about the orbital characteristics available. In the latter case, such information is available, and this allows it to linearize the problem and to solve it with standard procedures of applied mathematics. *Cum grano salis* one might say that first orbit determination is an *art*, whereas orbit improvement is *mathematical routine.*

Let us first comment the artistic task: If the force field is assumed to be known (in most cases one even uses the two-body approximation) the problem is reduced to determining six parameters, the (osculating) orbital elements, using the observations. First orbit determination can only succeed, if the number of unknowns can be reduced to only one or two parameters. The principle is explained in the case of determining a circular orbit using two astrometric positions, where the (originally) six-dimensional problem is reduced to one of dimension one.

In the general case, it is possible to reduce the problem to a two-dimensional problem, the topocentric distances corresponding to two astrometric positions being the remaining unknowns. The method is based on the numerical solution of a local boundary value problem as discussed in Chapter 7. The new method presented here is very robust, allowing it to investigate also delicate cases as, e.g., multiple solutions.

Classical orbit determination must be illustrated with standard and difficult examples. Program ORBDET, serving this purpose, allows it to determine first orbits of objects in the planetary system (minor planets, comets, NEO (Near Earth Objects), etc.), and of satellites or space debris. First orbits may be determined with a variety of methods in ORBDET, including the determination of circular and parabolic orbits. Except for the case of the circular orbit the basic method is the new method mentioned above. Sample observations of minor planets, comets, and artificial satellites are provided. Each orbit determination is concluded by an orbit improvement step, where the more important perturbations of the particular problem may be taken into account.

Most of the orbit determination procedures in use today are based on ideas due to Gauss, Laplace, and others. The historical reminiscences are discussed, but not considered for implementation. Often, the original recipes have been simply translated into a computer code and applied – from our point of view a totally unacceptable procedure. Our method outlined in the main text and implemented in program ORBDET is based on Gauss's brilliant insight that the formulation of the orbit determination problem as a boundary value problem (instead of an initial value problem) immediately reduces the number of parameters from six to two, and on the numerical solution of the associated (local) boundary value problem.

Not only angles, but also distances, distance differences, and other aspects of a satellite orbit may be observed. This naturally leads to much more general orbit determination problems in satellite geodesy. Usually, one may assume moreover that good approximations of the true orbits are known – meaning that "only" standard methods (based on linearizing a non-linear parameter estimation problem) are required to improve the orbits.

An observation of a celestial body does not only contain information concerning the position (and/or velocity) vector of the observed object, but also about the observer's position and motion. This aspect is widely exploited in satellite geodesy. Some of the general parameter estimation schemes and of the results achieved are briefly mentioned in Chapter 8, as well.

The chapter is concluded with two modern examples of "pure" orbit determination problems. One is related to SLR (Satellite Laser Ranging), the other to the determination of LEO (Low Earth Orbiter) orbits, where the LEO is equipped with a GPS receiver. The latter orbit determination problem is based on the LEO positions (and possibly position differences) as determined from the data of the spaceborne GPS receiver. Program SATORB may be used to determine these orbits. This latter application is attracting more and more attention because more and more LEOs are equipped with GPS receivers.

1.2 Part II: Applications

Rotation of Earth and Moon. Chapter II- 2 deals with all aspects of the three-body problem Earth-Moon-Sun. All developments and analyses are based on the corresponding equations of motion developed in Chapter 3; the illustrations, on the other hand, are based almost exclusively on the computer program ERDROT (see section 1.3).

In order to fully appreciate the general characteristics of Earth (and lunar) rotation, it is necessary to understand the orbital motion of the Moon in the first place. This is why the orbital motion of the Moon is analyzed before discussing the rotation of Earth and Moon.

The main properties of the rotation of Earth and Moon are reviewed afterwards under the assumption that both celestial bodies are rigid. Whereas the characteristics of Earth rotation are well known, the rotational properties of the Moon are usually only vaguely known outside a very limited group of specialists. Despite the fact that the structure of the equations is the same in both cases, there are noteworthy differences, some of which are discussed in this chapter. The analysis pattern is the same for the two bodies: The motions of the rotation axis in the body-fixed system and in the inertial system are established by computer simulations (where it is possible to selectively "turn off" the torques exerted by the respective perturbing bodies); the simulation results are then explained by approximate analytical solutions of the equations of motion. The simulations and the approximate analytic solutions are compared to the real motion of the Earth's and Moon's rotation poles. Many, but not all aspects are explained by the rigid-body approximation.

This insight logically leads to the discussion of the rotation of a non-rigid Earth. This discussion immediately leads in turn to very recent, current and possible future research topics. Initially, the "proofs" for the non-rigidity of the Earth are provided. This summary is based mainly on the Earth rotation series available from the IERS and from space geodetic analysis centers. Many aspects of Earth rotation may be explained by assuming the Earth to consist of a solid elastic body, which is slightly deformed by "external" forces. Only three of these forces need to be considered: (1) the centrifugal force due to the rotation of the Earth about its figure axis, (2) the differential centrifugal force due to the rotation of the Earth about an axis slightly differing from this figure axis, and (3) the tidal forces exerted by Sun and Moon (and planets). The resulting, time-dependent deformations of the Earth are small, which is why in a good approximation they may be derived from Hooke's law of elasticity. The elastic Earth model brings us one step closer to the actual rotation of the Earth: The difference between the Chandler and the Euler period as well as the observed bi-monthly and monthly LOD (Length of Day) variations can be explained now.

The elastic Earth model does not yet explain all features of the observed Earth rotation series. There are, e.g., strong annual and semi-annual variations in the real LOD series, which may *not* be attributed to the deformations of the solid Earth. Peculiar features also exist in the polar motion series. They are observed with space geodetic techniques because the observatories are attached to the solid Earth and therefore describe the rotation of this body (and not of the body formed by the solid Earth, the atmosphere and the oceans). Fortunately, meteorologists and oceanographers are capable of deriving the angular momentum of the atmosphere from their measurements: by comparing the series of AAM (Atmospheric Angular Momentum) emerging from the meteorological global pressure, temperature, and wind fields with the corresponding angular momentum time series of the solid Earth emerging from space geodesy, the "unexplained" features in the space geodetic observation series of Earth rotation are nowadays interpreted by the exchange of angular momentum between solid Earth, atmosphere and oceans – implying that the sum of the angular momenta of the solid Earth and of atmosphere and oceans is nearly constant.

Even after having modelled the Earth as a solid elastic body, partly covered by oceans and surrounded by the atmosphere, it is not yet possible to explain all features of the monitored Earth rotation. Decadal and secular motions in the observed Earth rotation series still await explanation. The explanation of these effects requires even more complex, multi-layer Earth-models, as, e.g., illustrated by Figure II- 2.55. The development of these complex Earth models is out of the scope of an introductory text. Fortunately, most of their features can already be seen in the simplest generalization, usually referred to as the Poincaré Earth model, consisting of a rigid mantle and a fluid core (see Figure II- 2.56). It is in particular possible to explain the terms FCN (Free Core Nutation) and NDFW (Nearly-Diurnal Free Wobble). The mathematical deliberations associated with the Poincaré model indicate the degree of complexity associated with the more advanced Earth models. It is expected that such models will be capable of interpreting the as yet unexplained features in the Earth rotation series – provided that Earth rotation is continuously monitored over very long time spans (centuries).

Artificial Earth Satellites. Chapter II- 3 deals with the orbital motion of artificial Earth satellites. Most illustrations of this chapter stem from program SATORB, which allows it (among others) to generate series of osculating and/or mean elements associated with particular satellite trajectories.

The perturbations of the orbits due to the oblate Earth, more precisely the perturbations due to the term C_{20} of the harmonic expansion of Earth's potential, are discussed first. The pattern of perturbations at first sight seems rather similar to the perturbations due to a third body: No long-period or secular perturbations in the semi-major axis and in the eccentricity, secular perturbations in the right ascension of the ascending node Ω and in the argument ω of perigee. There are, however, remarkable peculiarities of a certain practical relevance. The secular rates of the elements Ω (right ascension of ascending node) and ω (argument of perigee) are functions of the satellite's inclination i w.r.t. the Earth's equatorial plane. The perturbation patterns allow it to establish either sun-synchronous orbital planes or orbits with perigees residing in pre-defined latitudes.

The orbital characteristics are established by simulation techniques (using program SATORB), then explained with first-order general perturbation methods (based on simplified perturbative forces). Higher-order perturbations due to the C_{20}-term and the influence of the higher-order terms of the Earth's potential (which are about three orders of magnitude smaller than C_{20}) are studied subsequently. The attenuating influence of the Earth's oblateness term C_{20} on the perturbations due to the higher-order terms C_{ik} is discussed as well.

If a satellite's revolution period is commensurable with the sidereal revolution period of the Earth, some of the higher-order terms of Earth's potential may produce resonant perturbations, the amplitudes of which may become orders of magnitude larger than ordinary higher-order perturbations. Resonant perturbations are typically of very long periods (years to decades), and the amplitudes may dominate even those caused by the oblateness. Two types of resonances are discussed in more detail, the (1:1)-resonance of geostationary satellites and the (2:1)-resonance of GPS-satellites. In both cases the practical implications are considerable. In the case of GPS-satellites the problem is introduced by a heuristic study, due to my colleague Dr. Urs Hugentobler, which allows it to understand the key aspects of the problem without mathematical developments.

The rest of the chapter is devoted to the discussion of non-gravitational forces, in particular of drag and of solar radiation pressure. As usual in our treatment, the perturbation characteristics are first illustrated by computer simulations, then understood by first-order perturbation methods. Atmospheric drag causes a secular reduction of the semi-major axis (leading eventually to the decay of the satellite orbit) and a secular decrease of the eccentricity (rendering the decaying orbit more and more circular). Solar radiation pressure is (almost) a conservative force (the aspect is addressed explicitly), which (almost) excludes secular perturbations in the semi-major axis. Strong and long-period perturbations occur in the eccentricity, where the period is defined by the periodically changing position of the Sun w.r.t. the satellite's orbital plane.

The essential forces (and the corresponding perturbations) acting on (suffered by) high- and low-orbiting satellites conclude the chapter.

Evolution of the Planetary System. The application part concludes with Chapter II- 4 pretentiously entitled *evolution of the planetary system.* Three major issues are considered: (a) the orbital development of the outer system from Jupiter to Pluto over a time period of two million years (the past million years and the next million years – what makes sure that the illustrations in this chapter will not be outdated in the near future, (b) the orbital development of the complete system (with the exception of the "dwarfs" Mercury and Pluto), where only the development of the inner system from Venus to Mars is considered in detail, and (c) the orbital development of minor planets (mainly of those in the classical asteroid belt between Mars and Jupiter).

The illustrations have three sources, namely (a) computer simulations with program PLASYS, allowing it to numerically integrate any selection of planets of our planetary system with the inclusion of one body of negligible mass (e.g., a minor planet or a comet) with a user-defined set of initial orbital elements (definition in Chapter 2), (b) orbital elements obtained through the MPC (Minor Planet Center) in Cambridge, Mass., and (c) spectral analyses of the series of orbital elements (and functions thereof) performed by our program FOURIER.

By far the greatest part of the (mechanical) energy and the angular momentum of our planetary system is contained in the outer system. Jupiter and Saturn are the most massive planets in this subsystem. Computer simulations over relatively short time-spans (of 2000 years) and over the full span of two million years clearly show that even when including the entire outer system the development of the orbital elements of the two giant planets is dominated by the exchange of energy and angular momentum between them. The simulations and the associated spectra reveal much more information.

Venus and Earth are the two dominating masses of the inner system. They exchange energy and angular momentum (documented by the coupling between certain orbital elements) very much like Jupiter and Saturn in the outer system. They are strongly perturbed by the planets of the outer system (by Jupiter in particular). An analysis of the long-term development of the Earth's orbital elements (over half a million years) shows virtually "no long-period structure" for the semi-major axis, whereas the eccentricity varies between $e \approx 0$ and $e \approx 0.5$ (exactly like the orbital eccentricity of Venus).

Such variations might have an impact on the Earth's climate (annual variation of the "solar constant", potential asymmetry between summer- and winter-half-year). The eccentricity is, by the way, approaching a minimum around the year 35'000 A.D., which does not "promise" too much climate-relevant "action" in the near future – at least not from the astronomical point of view. The idea that the Earth's dramatic climatic changes in the past (ice-ages and warm periods) might at least in part be explained by the Earth's orbital motion is due to Milankovitch. Whether or not this correlation is significant cannot be firmly decided (at least not in this book). The long-term changes of the orbital characteristics (of the eccentricity, but also of the inclination of the Earth's orbital plane w.r.t. the so-called invariable plane) are, however, real, noteworthy and of respectable sizes.

Osculating orbital elements of more than $100'000$ minor planets are available through the MPC. This data set is inspected to gain some insight into the motion of these celestial objects at present. The classical belt of minor planets is located between Mars and Jupiter. Many objects belonging to the so-called Kuiper-belt are already known, today. Nevertheless, the emphasis in Chapter II- 4 is put on the classical belt of asteroids and on the explanation of (some aspects of) its structure. The histogram II- 4.43 of semi-major axes (or of the associated revolution periods) indicates that the Kirkwood gaps must (somehow) be explained by the commensurabilities of the revolution periods of the minor planets and of Jupiter. After the discussion of the observational basis, the analysis of the orbital motion of minor planets is performed in two steps:

- The development of the orbital elements of a "normal" planet is studied. This study includes the interpretation of the (amazingly clean) spectra of the minor planet's mean orbital elements. These results lead to the definition of the (well known) so-called proper elements. It is argued that today the definition of these proper elements should in principle be based on numerical analyses, rather than on analytical theories as, e.g., developed by Brouwer and Clemence [27]. A few numerical experiments indicate, however, that the results from the two approaches agree quite well.

- Minor planets in resonant motion with Jupiter are studied thereafter. The Hilda group ((3:2)-resonance) and the (3:1)-resonance are considered in particular. The Ljapunov characteristic exponent is defined as an excellent tool to identify chaotic motion. A very simple and practical method for its establishment (based on the solution of one variational equation associated with the minor planet's orbit) is provided in program PLASYS. The tools of numerical integration of the minor planet's orbit together with one or more variational equations associated with it, allow it to study and to illustrate the development of the orbital elements of minor planets in resonance zones. It is fascinating to see that the revolutionary numerical experiments performed by Jack Wisdom, in the 1980s, using the most advanced computer hardware available at that time, nowadays may be performed with standard PC (Personal Computer) equipment.

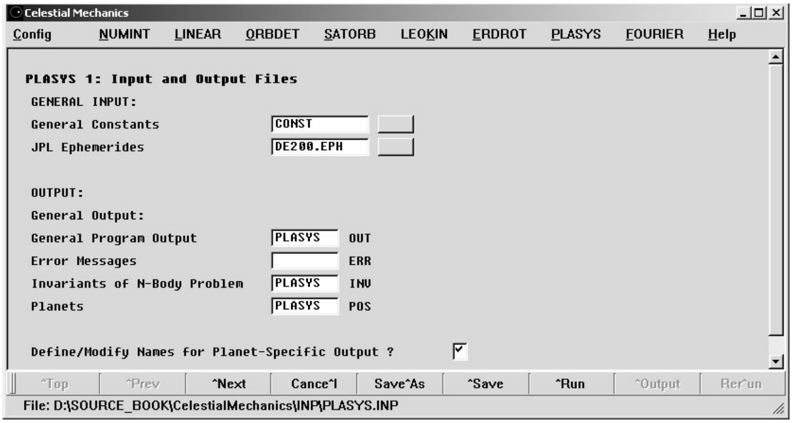
1.3 Part III: Program System

The program system, all the procedures, and all the data files necessary to install and to use it on PC-platforms or workstations equipped with a WINDOWS operating system are contained on the CDs accompanying both volumes of this work. The system consists of eight programs, which will be briefly characterized below. Detailed program and output descriptions are available in Part III, consisting of Chapters II- 5 to II- 11.

The program system is operated with the help of a menu-system. Figure 1.1 shows a typical panel – actually the panel after having activated the program system *Celestial Mechanics* and then the program PLASYS. The top line of each panel contains the buttons with the program names and the help-key offer real-time information when composing a problem.

The names of input- and output-files may be defined or altered in these panels and input options may be set or changed. By selecting ≪ Next Panel ≫ (bottom line), the next option/input panel of the same program are activated. If all options and file definitions are meeting the user's requirements, the program is activated by selecting ≪ Save and Run ≫ . For CPU (Central Processing Unit) intensive programs, the program informs the user about the remaining estimated CPU-requirements (in %).

The most recent general program output (containing statistical information concerning the corresponding program run and other characteristics) may be inspected by pressing the button ≪ Last Output ≫ . With the exception of LEOKIN all programs allow it to visualize some of the more specific output files using a specially developed graphical tool compatible with the menu-system. The output files may of course also be plotted by the program user with any graphical tool he is acquainted with. All the figures of this book



```
PLASYS 1: Input and Output Files
 GENERAL INPUT:
 General Constants              CONST
 JPL Ephemerides                DE200.EPH

 OUTPUT:
 General Output:
 General Program Output         PLASYS     OUT
 Error Messages                            ERR
 Invariants of N-Body Problem   PLASYS     INV
 Planets                        PLASYS     POS

 Define/Modify Names for Planet-Specific Output ?      ☑
```

File: D:\SOURCE_BOOK\CelestialMechanics\INP\PLASYS.INP

Fig. 1.1. Primary menu for program system *Celestial Mechanics, PLASYS*

illustrating computer output were, e.g., produced with the so-called "gnu"-graphics package. The gnu-version used here is also contained on the CD. The programs included in the package "Celestial Mechanics" are (in the sequence of the top line of Figure 1.1):

1. **NUMINT** is used in the first place to demonstrate or test the mutual benefits and/or deficiencies of different methods for numerical integration. Only two kinds of problems may be addressed, however: either the motion of a minor planet in the gravitational field of Sun and Jupiter (where the orbits of the latter two bodies are assumed to be circular) or the motion of a satellite in the field of an oblate Earth (only the terms C_{00} and C_{20} of the Earth's potential are assumed to be different from zero).

   The mass of Jupiter or the term C_{20} may be set to zero (in the respective program options), in which case a pure two-body problem is solved.

   When the orbit of a "minor planet" is integrated, this actually corresponds to a particular solution of the problème restreint. In this program mode it is also possible to generate the well known surfaces of zero velocity (Hill surfaces), as they are shown in Chapter 4.
2. **LINEAR** is a test program to demonstrate the power of collocation methods to solve linear initial- or boundary-value problems. The program user may select only a limited number of problems. He may test the impact of defining the collocation epochs in three different ways (equidistant, in the roots of the Legendre and the Chebyshev polynomials, respectively).
3. **SATORB** may either be used as a tool to generate satellite ephemerides (in which case the program user has to specify the initial osculating

elements), or as an orbit determination tool using *either* astrometric positions of satellites or space debris as observations *or* positions (and possibly position differences) as pseudo-observations. In the latter case SATORB is an ideal instrument to determine a purely dynamical or a reduced-dynamics orbit of a LEO. It may also be used to analyze the GPS and GLONASS ephemerides routinely produced by the IGS (International GPS Service).

The orbit model can be defined by the user, who may, e.g.,

- select the degree and the order for the development of the Earth's gravity potential,
- decide whether or not to include relativistic corrections,
- decide whether or not to include the direct gravitational perturbations due to the Moon and the Sun,
- define the models for drag and radiation pressure, and
- decide whether or not to include the perturbations due to the solid Earth and ocean tides.

Unnecessary to point out that this program was extensively used to illustrate Chapter II- 3.

When using the program for orbit determination the parameter space (naturally) contains the initial osculating elements, a user-defined selection of dynamical parameters, and possibly so-called pseudo-stochastic pulses (see Chapter 8).

Programs ORBDET and SATORB were used to illustrate the algorithms presented in Chapter 8.

4. **LEOKIN** may be used to generate a file with positions and position differences of a LEO equipped with a spaceborne GPS-receiver. This output file is subsequently used by program SATORB for LEO orbit determination. Apart from the observations in the standard RINEX (Receiver Independent Exchange Format), the program needs to know the orbit and clock information stemming from the IGS.

5. **ORBDET** allows it to determine the (first) orbits of minor planets, comets, artificial Earth satellites, and space debris from a series of astrometric positions. No initial knowledge of the orbit is required, but at least two observations must lie rather close together in time (time interval between the two observations should be significantly shorter than the revolution period of the object considered).

   The most important perturbations (planetary perturbations in the case of minor planets and comets, gravitational perturbations due to Moon, Sun, and oblateness of the Earth (term C_{20}) in the case of satellite motion) are included in the final step of the orbit determination. ORBDET is the only interactive program of the entire package.

The program writes the final estimate of the initial orbital element into a file, which may in turn be used subsequently to define the approximate initial orbit, when the same observations are used for orbit determination in program SATORB.

6. **ERDROT** offers four principal options:

   - It may be used to study Earth rotation, assuming that the geocentric orbits of Moon and Sun are known. Optionally, the torques exerted by Moon and Sun may be set to zero.
   - It may be used to study the rotation of the Moon, assuming that the geocentric orbits of Moon and Sun are known. Optionally, the torques exerted by Earth and Sun may be set to zero.
   - The N-body problem Sun, Earth, Moon, plus a selectable list of (other) planets may be studied and solved.
   - The program may be used to study the correlation between the angular momenta of the solid Earth (as produced by the IGS or its institutions) and the atmospheric angular momenta as distributed by the IERS (International Earth Rotation and Reference Systems Service).

   This program is extensively used in Chapter II- 2.

7. **PLASYS** numerically integrates (a subset of) our planetary system starting either from initial state vectors taken over from the JPL (Jet Propulsion Laboratory) DE200 (Development Ephemeris 200), or using the approximation found in [72]. A minor planet with user-defined initial osculating elements may be included in the integration, as well. In this case it is also possible to integrate up to six variational equations simultaneously with the primary equations pertaining to the minor planet. Program PLASYS is extensively used in Chapter II- 4.

8. **FOURIER** is used to spectrally analyze data provided in tabular form in an input file. The program is named in honour of Jean Baptiste Joseph Fourier (1768–1830), the pioneer of harmonic analysis. In our treatment Fourier analysis is considered as a mathematical tool, which should be generally known. Should this assumption not be (entirely) true, the readers are invited to read the theory provided in Chapter II- 11, where Fourier analysis is developed starting from the method of least squares. As a matter of fact it is possible to analyze a data set using

   - *either* the *least squares* technique – in which case the spacing between subsequent data points may be arbitrary,
   - *or* the *classical Fourier analysis*, which is orders of magnitude more efficient than least squares (but requires equal spacing between observations), and where *all* data points are used,
   - *or* FFT (Fast Fourier Transformation), which is in turn orders of magnitude more efficient than the classical Fourier technique, but where

usually the number of data points should be a power of 2 (otherwise a loss of data may occur).

In the FFT-mode the program user is invited to define the decomposition level (maximum power of 2 for the decomposition), which affects the efficiency, but minimizes (controls) loss of data. The general program output contains the information concerning the data loss.

The program may very well be used to demonstrate the efficiency ratio of the three techniques, which should produce identical results. FOURIER is a pure service program.

The computer programs of Part III are used throughout the two volumes of our work. It is considered a minimum set ("starter's kit") of programs that should be available to students entering into the field of Astrodynamics, in particular into one of the applications treated in Part II of this work. The programs NUMINT, LINEAR, and PLASYS are also excellent tools to study the methods of numerical integration.

2. Historical Background

Celestial Mechanics deals with the orbital and rotational motion of celestial bodies, e.g., the dynamics of stellar systems, the motion of stars within galaxies, the dynamics of planetary systems. In this book we focus on the orbital motion of planets, minor planets, and comets in our solar system, on the orbital motion of artificial satellites around the Earth, on the orbital and rotational motion of Earth and Moon, and on the development of the planetary system. In section 2.1 we focus on aspects of the history of Celestial Mechanics related to the planetary system, in section 2.2 on the same aspects related to the Earth-near space and to the aspects of the modern realizations of the celestial and terrestrial reference systems.

2.1 Milestones in the History of Celestial Mechanics of the Planetary System

We will consider two aspects in this overview: those related to eminent scientists, summarized in Table 2.1, and those related to important discoveries in the planetary system, summarized in Table 2.3.

The history of Celestial Mechanics should start with the first attempts to observe and predict the apparent motions of the Moon, the Sun, and the planets w.r.t. the celestial sphere of fixed stars. Our story, however, begins in the 16$^{\text{th}}$ century with the epoch of Tycho Brahe (1546–1601) and Johannes Kepler (1571–1630). Towards the end of the 16th century Brahe had set new standards in astronomical observation techniques. First in Denmark (1576–1597), then in Prague (1599–1601), he and his collaborators observed the positions of the planets and the Sun w.r.t. the zodiacal stars with an accuracy of about 1–2 arcminutes ($'$). Loosely speaking, the position of a celestial body is the direction from the observer to the observed object at a certain observation epoch. The astronomical position is characterized by a unit vector (which in turn may be specified by two angles) and the corresponding observation time. The accuracy achieved by Brahe was one of the best in the pre-telescope era, close to the best that could be obtained with only mechanical observation techniques.

In view of the accuracy of the position measurements, the observation epoch had to be accurate to a few seconds of time. The best time scale then available was given by the Earth's rotation (sidereal or solar time), defined in turn by observations of the stars and the sun. This is why Brahe also needed to observe stars in his comprehensive observation program. For interpolation of time, mechanical clocks with an accuracy of few seconds over, let us say, one day already were available. These few remarks show that Brahe did not only "observe a few planets", but that he and his team accomplished a rather comprehensive observational survey in astronomy towards the end of the 16th century, lasting for a time period of about a quarter of a century. Brahe's program should be compared with current international observation programs in astronomy, organized, e.g., by the IERS (International Earth Rotation and Reference Systems Service). Brahe's main result was a very consistent, long, and complete set of positions for the sun and the planets.

Around 1600 Kepler was *Landschaftsmathematiker* (state surveyor) in Graz. His inclination towards astronomy was documented by his work *Mysterium cosmographicum*, where he tried to relate the radii of the classical planets to the five regular polyhedra. Although Brahe was not so impressed by Kepler's work he invited Kepler as a co-worker to Prague. Tycho certainly hoped that Kepler would help him to further develop his own model of the planetary system, a mixture of the systems due to Nicholas Copernicus (1473–1543) and Claudius Ptolemaeus (ca. 100–170). Kepler had different ideas, however. We know from history that Kepler tried to find a physical law governing the planetary motions. He used the law of areas and introduced ellipse-shaped orbits *to reduce the calculations* when processing Tycho's time series of the positions of the sun and of the planet Mars. The steps eventually leading to the so-called Kepler's first two laws are documented in the *Astronomia nova* (which appeared in 1609) and in the correspondence between Kepler and his teacher Michael Mästlin (1550–1631) as well as Kepler's "rival" Longomontanus (1562–1647). However, the first two "laws" may not be found in the *Astronomia nova* because Kepler failed to confirm them by theory and observation. The third law was published only in the *Harmonice mundi* of 1619.

Let us include Kepler's laws in modern language:

1. The orbit of each planet around the sun is an ellipse with the Sun at one of its foci.
2. Each planet revolves so that the line joining it to the Sun sweeps out equal areas in equal (intervals of) time. (*Law of areas*).
3. The periods of any two planets are proportional to the 3/2 powers of their mean distances from the sun.

The first law implies that planetary motion is taking place in orbital planes, characterized, e.g., by two angles (in the planetary system, e.g., Ω the eclip-

Table 2.1. Highlights in the history of Celestial Mechanics

Name	Year	Event
T. Brahe	1600	Long and accurate (about 1–2 arcmin) time series of observations of planets and sun
J. Kepler	1609	*Astronomia nova* published
I. Newton	1687	Publication of the *Philosophiae naturalis principia mathematica*
L. Euler	1749	Modern version of eqns. of motion published in *Recherches sur le mouvement des corps célestes en général*
J. L. Lagrange	1779	Introduction of perturbing function
P. S. Laplace	1798	*Traité de mécanique céleste* published
C. F. Gauss	1801	Orbit determination for planetoid *Ceres*
U. J. J. Leverrier	1846	Analysis of perturbation of Uranus leads to detection of Neptune
G. W. Hill	1878	*Researches in the lunar theory.* Periodic *variational orbit* for Moon, establishment of Hill's eqns. of motion, etc.
F. Tissérand	1889	*Traité de mécanique céleste* contains, e.g., criterion on identity of comets
H. Poincaré	1889	The work *Sur le problème des trois corps et les équations de dynamique* wins the price of the Swedish king
	1892	*Les méthodes nouvelles de la mécanique céleste* initiates the research on dynamical systems
S. Newcomb	1900	Basis for production of ephemerides (almanacs) in planetary system
A. Einstein	1915	Theory of General Relativity as new fundament of Celestial Mechanics
K. Hirayama	1918	Discussion of families of minor planets
A. N. Kolmogorov, V. Arnold, J. Moser	1963	Not all series developments of three body problem are diverging
J. Wisdom	1987	In-depth analysis of chaotic movement in planetary system
A. E. Roy	1988	Project Longstop, integration of planetary system over 100 million years

tical longitude of the ascending node, and the inclination i with respect to the ecliptic). The first law furthermore implies that there must be a point of closest approach to the sun, the so-called perihelion, characterized by the angle ω, the argument of perigee (angle between ascending node and perigee). Size and shape of the ellipse are defined by the semi-major axis a and the eccentricity e. If we add the time T_0 of perihelion passage (or simply time of perihelion) to these five geometrical elements we obtain a set of six orbital elements a, e, i, Ω, ω, and T_0, which we still use today to characterize the orbits of celestial bodies. These orbital elements are illustrated in Figure 2.1. *Cum grano salis* we may say that we owe Kepler the orbit parametrization still in use today. We know that Kepler's laws are only correct for the "pure"

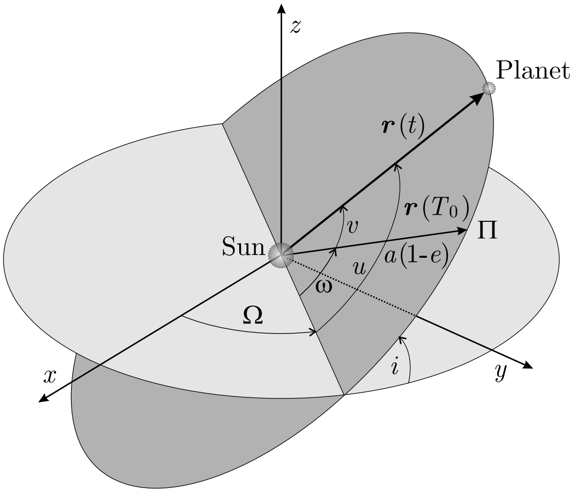


Fig. 2.1. Keplerian elements

two-body problem, i.e., in the absence of perturbing forces due to other planets. It is always possible, however, to view a trajectory as a time evolution of the orbital elements mentioned.

The law of areas, illustrated in Figure 2.2, implies that the velocity of a planet near perihelion exceeds that near aphelion. The law of areas is used to compute the *true anomaly* v (the angle between the direction to the perihelion and the current position, as seen from the Sun) as a function of time (see Chapter 3). Kepler's first two laws allow it to compute position and velocity of a celestial body at any given instant of time t. Kepler's laws therefore allow it to calculate the ephemerides of the planets in a very simple way.

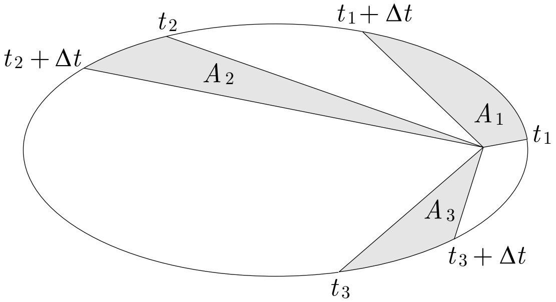
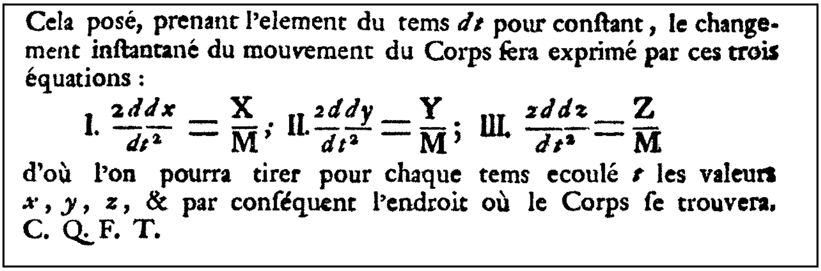


Fig. 2.2. Kepler's law of areas

In 1687 Sir Isaac Newton (1643–1727) published his *Philosophiae naturalis principia mathematica.* Newton's famous work [83] contains the fundaments of mechanics as they were understood in the 17th century. Because of their tremendous impact on the development of physics in general and Celestial Mechanics in particular we quote some of the important definitions in Table 2.2 using the recent English translation by Cohen and Whitman [84].

The chapter *Definitiones* (Definitions) deals with questions like what is matter?; what is linear momentum (as we call it today)?; what is a force?; what is absolute time?; what is absolute space? The second chapter *Axiomata sive Leges motus* (Axioms, the Laws of Motion) develops the fundaments of dynamics. Three so-called Books follow this introductory part. The first two are entitled *De motu corporum* (about the motion of bodies), the third *De mundi systemate* (about the system of the world). The quotes in Table 2.2 stem from the introduction and this third Book, which develops the law of universal gravitation. The first definition in principle declares mass as the product of volume and density, the second the (linear) momentum as the product of mass and velocity in “absolute space”. The third defines the inertia of a body, and the fourth introduces the concept of force as the only reason for a body to change its state of motion. It is interesting that Newton discusses space, time, “place”, and motion only after these definitions (which already require an understanding of these notions) in a section called *scholium.* Reading the second law, we immediately recognize the equations of motion. Newton's wording is even general enough to derive from this law the equation of motion of a rocket (with variable mass). This understanding, however, does not reflect the historical truth. The equations of motion, as we still use them today, are due to Leonhard Euler (1707–1783), who stated them in the form still used today in his work *Recherches sur le mouvement des corps célestes en général* published in 1749 (see Figure 2.3), more than sixty years after the publication of Newton's *Principia.* In 1750 he recognized that these equations are valid for any mass element and thus define a “new” mechanical principle [34]. Book 3 of Newton's *Principia*, *De mundi system-*



Cela poſé, prenant l'element du tems dt pour conſtant, le changement inſtantané du mouvement du Corps ſera exprimé par ces trois équations :

$$\text{I. } \frac{2ddx}{dt^2} = \frac{X}{M}; \quad \text{II. } \frac{2ddy}{dt^2} = \frac{Y}{M}; \quad \text{III. } \frac{2ddz}{dt^2} = \frac{Z}{M}$$

d'où l'on pourra tirer pour chaque tems ecoulé t les valeurs x, y, z, & par conſéquent l'endroit où le Corps ſe trouvera. C. Q. F. T.

Fig. 2.3. The equations of motion in Euler's work of 1749

Table 2.2. Quotes from Newton's *Principia*

Part	Statement
Definitions	*Def. 1:* Quantity of matter is a measure of matter that arises from its density and its volume jointly. *Def. 2:* Quantity of motion is a measure of motion that arises from the velocity and the quantity of matter jointly. *Def. 3:* Inherent force of matter is the power of resisting by which every body, so far as it is able, perseveres in its state either of resting or of moving uniformly straight forward. *Def. 4:* Impressed force is the action exerted on a body to change its state either of resting or of moving uniformly straight forward.
Scholium	*1:* Absolute, true, and mathematical time, in and of itself and of its own nature, without reference to anything external, flows uniformly and by another name is called duration. ... *2:* Absolute space, of its own nature and without reference to anything external, always remains homogeneous and immovable. ... *3:* Place is the part of space that a body occupies, ... *4:* Absolute motion is the change of position of a body from one absolute place to another; ...
Axioms	*Law 1:* Every body perseveres in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by forces impressed. *Law 2:* A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed. *Law 3:* To any action there is always an opposite and equal reaction; in other words, the actions of two bodies upon each other are always equal and always opposite in direction. *Corollary 1:* A body acted on by [two] forces acting jointly describes the diagonal of a parallelogram in the same time in which it would describe the sides if the forces were acting separately.
Book 3	*Theorem 7:* Gravity exists in all bodies universally and is proportional to the quantity of matter in each. *Theorem 8:* If two globes gravitate toward each other, and their matter is homogeneous on all sides in regions that are equally distant from their centers, then the weight of either globe towards the other will be inversely as the square of the distance between the centers.

ate, deals almost uniquely with the law of universal gravitation. After giving rules for professional work in natural philosophy (which still should be observed today), he states in a chapter called *phenomena* that the orbits of the satellites of Jupiter and Saturn in their orbit around their planets, the orbits of the five (classical) planets and that of the Earth around the Sun, and the orbit of the Moon around the Earth are all in agreement with Kepler's laws of planetary motion (after suitable generalization). In the chapter *propositions* Newton postulates (and proves these propositions making use of results from

Books 1 and 2) that the force responsible for all the above mentioned orbital motions is universal gravitation. He then concludes that the theorems 7 and 8 in Table 2.2 must hold. In mathematical terms, the law of gravitation states that two bodies (of the type specified in theorem 8) separated by a distance r attract each other by a force

$$\boldsymbol{f} = G \, \frac{m \, M}{r^2} \, \boldsymbol{e} \quad , \tag{2.1}$$

where G is the *gravitational constant*, m and M are the masses of the two bodies, and $\boldsymbol{e}$ is the unit vector from the attracted to the attracting body. Experts in the field claim that Robert Hooke (1635–1702) initially played a central role in the development of the law of gravitation, particularly by recognizing the importance of *centripetal* forces and by the idea of decomposing the orbital trajectory into a tangential and a radial component, which points to the "attractive" central body, and by explaining the change of the tangential component by "pulses" (instantaneous velocity changes) in the radial direction. It was, however, Newton's achievement to explain Kepler's laws (more or less rigorously) by an inverse-square-law (the so-called *direct* and *inverse problem*), to recognize gravitation as a property of matter, and thus to postulate the universal law of gravitation.

Newton's discussion in Book 3 of the *Principia* is much more general than what is reflected in Table 2.2. He states, e.g., that the gravitational attraction on a body exerted by an arbitrary mass distribution may be computed as the superposition of the gravitational attractions of the mass elements of the mass distribution. We will use this principle in Section 3.3 to derive the equations of motion for the three-body problem Earth-Moon-Sun.

Perturbation theory is an essential instrument in Celestial Mechanics of the planetary system. We make use of the fact that the motion in the ellipse is a good approximation of the actual motion. Loosely speaking, we may deal only with the difference "solution of the actual problem minus the corresponding elliptical solution". This difference may be rendered zero at one particular epoch and it is small in absolute value in the vicinity of this epoch. The goal of analytical perturbation theory is the approximative solution of the equations of motion in terms of known base functions (usually trigonometric series) which may be easily integrated.

Euler, Alexis-Claude Clairaut (1713–1765), Jean Le Rond d'Alembert (1717–1783), and Joseph Louis de Lagrange (1736–1813), were pioneers of perturbation theory. The introduction of the (scalar) perturbation function, e.g., is due to Lagrange. In his *Méchanique analitique* (1788) we find an encompassing compilation of analytical methods (the keywords being Lagrange brackets, Lagrangian perturbation equations). Pierre Simon de Laplace (1749–1827) was an accomplished master of Celestial Mechanics. He left behind the five volumes of his *Traité de mécanique céleste*. He seems to have introduced the notion *mécanique céleste*, Celestial Mechanics. Using his methods of perturbation theory he was able to explain the observed, seemingly secular perturbations of Jupiter and Saturn (great inequality). His success in explaining this and other delicate phenomena in the solar system probably led him to his belief of predictability of all phenomena (not only in the planetary system) over arbitrary time spans, provided the initial state of the system was known with sufficient accuracy. In this context we still speak of the *Laplacian demon.* Laplace was convinced of the stability of the solar system and he thought that he had proved this statement. His conclusions were, however, based on a rather uncritical use of series expansions.

Urbain Jean Joseph Leverrier (1811–1877) and later on Henri Poincaré (1854–1912) had serious doubts concerning the stability of the planetary system, or, to be more precise, concerning the validity of the proofs due to Laplace. With the advancement of mathematical analysis and with the improvement of the methods of analytical mechanics it became possible to show that some of the series expansions which were previously thought to be convergent actually were not. Poincaré showed that even in the restricted three-body problem (see below) there were cases, where two orbits which are infinitesimally close at a time t_0 will deviate from each other exponentially as a function of time. This is exactly what we understand today by the term *deterministic chaos.* Poincaré is the father of the theory of dynamical systems. His *Méthodes nouvelles de la mécanique céleste* are worth to be red even today. With the new English edition [86] it is possible to fully appreciate his contribution to Celestial Mechanics in particular and to the analysis of dynamical systems in general. The three volumes entitled *Integral invariants and asymptotic properties of certain solutions, Approximations by series*, and *Periodic and asymptotic solutions* demonstrate his interests quite well. For relaxation and entertainment we also refer to [85] in this context.

The circumstance that the so-called two-body problem has simple “analytic” solutions but that the general three-body problem cannot be solved in closed form led to many attempts to reduce the latter problem in such a way that an analytical solution becomes feasible. The *problema restrictum* was studied for the first time by Euler in 1766, then by Lagrange as *problème restreint.* The well known five stationary solutions of the restricted problem are partly due to Euler and to Lagrange. George William Hill (1838–1914) developed his lunar theory by studying the actual motion relative to a periodic solution of the three-body problem Earth-Moon-Sun. It is ironical that Poincaré, who wanted to prove the stability of the restricted problem by representing an actual orbit as infinitesimally close to a suitable periodic solution (which is stable by definition), found that for some of the resonant motions the opposite was true. It often happens that a problem, which is simplified in the attempt to find simple approximations, is no longer strongly related to the original problem. The (restricted) three-body problem, however, proved to be most stimulating for the advancement of Celestial Mechanics.

The analytical methods of perturbation theory were of greatest importance and impact for practical work in astronomy, in particular for the production of the ephemerides in the solar system. The *Nautical Almanac and Astronomical Ephemeris*, predecessor of today's *Astronomical Almanac*, and the *The American Ephemeris and Nautical Almanac* (see [107]), were based on Simon Newcomb's (1835–1909) Tables of Sun, Moon and Planets. Today, the ephemerides are based on the technique of numerical integration (see [107]).

Perturbation theory has already brought us deeply into the 20th century. Let us return now to the 18th century to comment some of the exciting discoveries in the planetary system since the invention of the telescope. Some of the highlights are given in Table 2.3.

On March 13, 1781 Uranus, the seventh planet of our solar system, was discovered by John Frederick William Herschel (1738–1822). Herschel never agreed that he discovered Uranus by chance, but that he owed this success to his systematic survey of the skies. This is of course true, but it is also true that Herschel's motivation for his systematic optical survey was not the search for planets. The discovery of Uranus must have been an epoch-making event in the 18th century. At once, the "god-given" number of the six classical planets Mercury, Venus, Earth, Mars, Jupiter and Saturn had changed. Strangely enough, the semi-major axis of Uranus' orbit seemingly confirmed the empirical rule set up by Johann Daniel Titius (1747–1826) in 1766 relating the semi-major axes of the planetary orbits by a geometrical series (today written as):

Table 2.3. Discoveries in the planetary system

Year	Discovery
1781	*Uranus* by Herschel
1801	First minor planet, *Ceres*, by Piazzi
1846	*Neptune* by Galle, based on predictions by Leverrier (and Adams)
1930	*Pluto* by Clyde William Tombaugh (1906–1997) at Lowell Observatory
1977	*Chiron*, first minor planet with aphelion far beyond Jupiter (Kuiper belt), discovered by Kowal

$$a = 0.4 + 0.3 \cdot 2^n \,\mathrm{AU}\ , \quad n = -\infty, 0, 1, 2, \dots \quad .$$

Johann Elert Bode (1729–1796) was making this rule publicly known in 1772. It is a bit strange that there are no numbers between $-\infty$ and 0 in this rule, but Table 2.4 illustrates how well the rule holds in the planetary system. We should keep in mind that the rule was set up prior to the discovery of Uranus, which was interpreted as a strong evidence that this rule was a new law of nature which had to be explained by the scientists! It was troubling that, according to this rule, there was no planet between Mars and Jupiter corresponding to a semi-major axis of $a \approx 2.7$ AU. The belief in the Titius-Bode law was so strong that a coordinated search in Germany for the new planet between Mars and Jupiter was initiated. The organized search for it was not successful. On the other hand, Giuseppe Piazzi (1746–1826) discovered a faint new planet in the evening of the New Year's day 1801 – a perfect way for an astronomer to commence a new century. Ceres, as the new "planet" was called, proved to be the biggest object in a long series of minor planets discovered in the following years. Piazzi could observe Ceres only 19 times in January and early February 1801 during a period of 42 days following the discovery. The time period of about 40 days is rather short when compared to the revolution period of the planet of about four years. There was a certain danger that the newly discovered planet would again be lost "forever". In 1801 there were no methods available allowing it to derive the orbital elements of a celestial object in the solar system from a short time series of direction observations.

Table 2.4. Titius-Bode rule

Planet	True axis a	Axis a	Rev. Period
	[AU]	Titius-Bode [AU]	[Years]
Mercury	0.39	0.40	0.24
Venus	0.72	0.70	0.62
Earth	1.00	1.00	1.00
Mars	1.52	1.60	1.88
Minor Planets	~ 2.7	2.8	~ 4.44
Jupiter	5.20	5.2	11.86
Saturn	9.54	10.0	29.46
Uranus	19.19	19.6	84.02
Neptune	30.06	—	164.79
Pluto	39.53	38.8	249.17

The discovery of Ceres in 1801 and of other minor planets soon thereafter initiated a new branch in Celestial Mechanics, that of *first orbit determination.* Two scientists have to be mentioned in this context, namely the famous German mathematician Carl Friedrich Gauss (1777–1855) (see Table 2.1) and the French specialist in Celestial Mechanics, Laplace (see also Table 2.1). Their concepts of so-called first orbit determination both are most attractive from the mathematical point of view and they are quite different in nature. We will address the topic of first orbit determination in detail in Chapter 8.

The algorithm provided by Gauss proved to be very robust and most successful; it was used by many generations of astronomers and it is still used in modified form in the computer age. Gauss became famous in the greater scientific community (also) thanks to his successful orbit determination for Ceres, which allowed a safe re-discovery of the celestial body on December 7, 1801. It was impressive that the semi-major axis $a \approx 2.77$ AU of its orbit did fit quite well into the Titius-Bode scheme in Table 2.4.

The next important event in Table 2.3 is the discovery of Neptune by Johann Gottfried Galle (1812–1910) in 1846 in Berlin. This discovery was a triumph of perturbation theory. In the first half of the 19th century the experts in the field became aware of (periodic) orbit perturbations of Uranus that could not be explained by the gravitational perturbations from the known planets. Leverrier and independently John Cough Adams (1819–1892) tried to explain the perturbations by a new outer planet. The inverse task of perturbation theory (determination of orbit and mass of a perturbing body based on the orbital behavior of a known planet) is a most delicate problem. It cannot be addressed without adopting simplifying assumptions (e.g., circular orbit of perturbing body with known semi-major axis, which are then iteratively improved). It is interesting and convincing that both, the analyses by Leverrier and Adams, led to similar results. Based on the computation by Leverrier Galle found the new planet, subsequently called Neptune, only $4'$ away from the predicted position.

Leverrier wanted to repeat his success. By a very careful application of perturbation theory, taking into account the perturbing effects of all known planets he convincingly proved that about $43''$ per century of the secular perihelion motion of Mercury could not be explained. This part of Leverrier's work is a masterpiece. Less convincing is the second half of the story: Leverrier tried to explain this effect by a planet called *Volcano* with an orbit inside that of Mercury. To make a long story short: Table 2.3 documents that Volcano was never discovered. Other attempts to explain the excess of rotation of Mercury's perihelion (which was undoubtedly real), e.g., by a ring of dust around the Sun, also failed. Long after the establishment of Mercury's excess perihelion motion Albert Einstein's (1879–1955) general theory of relativity eventually explained the phenomenon – ironically enough as a consequence of gravitation (in the framework of general relativity), an explanation which was ruled out as a possible explanation by most experts early in the 20^{st} century. The story is exciting and it is well documented in [91].

Let us once more address the second event in Table 2.3. Shortly after the discovery of Ceres other minor planets were discovered, in particular Pallas in 1802, Juno in 1804, and Vesta in 1807. By 1850 about 150 minor planets or planetoids were known. Today, for more than 10000 of these objects excellent orbit information is available, e.g., through the *MPC (Minor Planet Center)* in Cambridge, USA, of the IAU (International Astronomical Union). The use of photography, and later on in the 20^{th} century the use of CCD (Charge Coupled Devices), rendered the discovery of fainter and fainter objects much easier. Today, thousands of new minor planets are discovered every month.

A histogram of the semi-major axes or the corresponding revolution periods reveals that the distribution between Mars and Jupiter is far from regular. There are maxima and gaps which cannot be explained by natural fluctuations. Based on a relatively modest sample of about fifty planets Daniel Kirkwood (1814–1895) was the first to describe such gaps in the last century, which is why they are called today Kirkwood gaps. Because many gaps and maxima correspond to orbits for which the fraction of the revolution periods of Jupiter and minor planets is a ratio of small integer numbers, it is fair to guess that the observed histogram may be explained by gravity alone. Strong evidence emerges from numerical experiments as performed, e.g., by Jack Wisdom [131], that Newtonian mechanics, combined with considerations of the probabilities of encounters are sufficient to explain the gaps.

Is there more to say concerning the structure of the population of minor planets? At the beginning of the 20th century Kiyotsugu Hirayama (1874–1943) showed that there are families of minor planets with similar orbit characteristics (semi-major axis, eccentricity, inclination), which might have been created by fragmentation from one proto-planetoid.

Fast electronic computers are essential tools for modern research addressing the structure and evolution of the planetary system. They allow it to study the evolution of the entire planetary system (including minor planets) over millions of years. We will again address this topic in Chapter II- 4. Let us mention that progress was also made in the theoretical domain. In the second half of the century (see Table 2.1) it was possible to demonstrate in the framework of the KAM theory (Kolmogoroff, Arnold, Moser) that some series developments in Celestial Mechanics are convergent after all. Such results are of considerable importance for the stability of the planetary system.

In the night of 18–19 October 1977 Charles T. Kowal discovered Chiron, the first minor planet between Saturn and Uranus, using the Schmidt Camera of Palomar Observatory (see Table 2.3). Chiron has a very interesting orbit lying almost entirely between Saturn and Uranus, with close encounters with the two planets making a long-time prediction of its orbit a delicate issue. The discovery seemed to indicate that minor planets are not confined to the region between Mars and Jupiter, but that they are common in the outer planetary system as well. More than 100 of these objects in the so-called Kuiper belt, named after Gerard Peter Kuiper (1905–1973), were discovered up to now.

2.2 The Advent of Space Geodesy[1]

The Earth has one natural satellite, namely the Moon, and hundreds of artificial satellites. In view of the fact that the first artificial Earth satellite, Sputnik-I, was only launched on October 4 of the International Geophysical Year 1957 this statement is quite amazing.

The Moon is of particular interest in Celestial Mechanics. Due to its closeness to the optical astronomical observatories on the surface of the Earth, not only its orbital motion may be studied with great accuracy, but also its rotational motion. The study of the orbital motion of the Moon led to the detection of irregularities in the time scales related to the rotation of the Earth and, as a consequence in the 1960s, to the introduction of *ephemeris time* as a more uniform *absolute* Newtonian time based on the orbital motion of the Moon (and of the planets). After the deployment of retro-reflectors on the surface of the Moon in the second half of the $20^{\rm th}$ century, *LLR (Lunar Laser Ranging)* allowed it to monitor the orbital (and rotational) motion of the Moon with the unprecedented accuracy of (few) cms. Due to the excellent observability and due to the Moon's almost perfect insensitivity w.r.t. non-gravitational forces (see Table II- 3.4 in Chapter II- 3), the orbit of the Moon proved to be an ideal test object for the theory of general relativity (more details will be provided in Chapter 3).

The three-body problem Earth-Moon-Sun is probably the best studied *real* three-body problem in Celestial Mechanics : A profound analysis not only gives insight into the orbital motion of the three bodies, but also into the rotation of both, the Earth and the Moon. The equations of motion for this problem, considering Earth and Moon as finite bodies, are set up in Chapter 3.3, the rotational motion of Earth and Moon is studied in considerable detail in Chapter II- 2.

Definitions and Principles of Space Geodesy. *Geodesy* studies the size and the figure of the Earth including the determination of the Earth's gravity field. *Geodetic astronomy* is that part of astronomy dealing with the definition and realization of a terrestrial and a celestial reference frame. *Space geodesy* addresses those aspects of geodesy and geodetic astronomy which are studied by using natural or artificial celestial bodies as observed objects or as observing platforms. In the older literature the term *Cosmic geodesy* is sometimes used as a synonym. Space geodesy is thus defined through the observation techniques, also referred to as *space geodetic techniques (methods).*

Space geodesy evolved rapidly in the second half of the twentieth century. In the space age it became possible to deploy and use artificial satellites *either* to study size and figure of the Earth from space *or* to observe them as targets

[1] based on the contribution *Space geodesy* by the author to the Encyclopedia of Astronomy [79]

from the surface of the Earth. The use of artificial Earth satellites for geodetic purposes is also referred to as *satellite geodesy.* The second essential development consists of the *VLBI (Very Long Baseline Interferometry)* technique as a new tool to realize an extraordinarily accurate and inertial reference system (called inertial reference frame) and to monitor Earth rotation using Quasars (Quasistellar Radio Sources).

Today, space geodetic techniques are the primary tools to study size, figure and deformation of the Earth, and its motion as a finite body w.r.t. the inertial reference system. Space geodetic techniques thus are the fundamental tools for geodesy, geodetic astronomy, and geodynamics.

Each space geodetic observation contains information about the position (and motion) of the observed object and the observer. Therefore, these observations also contain information concerning the transformation between the terrestrial and the inertial systems. The Earth orientation parameters, i.e., polar motion, UT1 (the time determined by the rotating Earth), precession and nutation define this transformation.

The Role of the Earth's Atmosphere. The signals of the observed or observing celestial bodies have to travel through the Earth's atmosphere. This changes the paths and the travel times of the signals. These effects are referred to as *refraction effects.* Refraction is usually considered a nuisance in astronomy, geodesy and geodynamics. In recent years refraction effects are more and more understood as a primary source of information for atmosphere science and are systematically monitored by space geodetic methods.

Whether the atmosphere related signal is useful depends on the wavelengths of the analyzed signals. If we measure, e.g., distances or distance differences to satellites using optical signals, refraction effects may be computed with sub-centimeter accuracy using pressure, temperature and humidity registrations at the observing sites. This implies that Laser ranging is not very useful for atmosphere monitoring. This fact may, however, also be formulated positively: Laser observations are well suited for calibrating other space geodetic techniques, which are more prone to atmospheric effects.

For microwave techniques like the Doppler systems, the GPS (Global Positioning System), the VLBI, one has to distinguish between ionospheric refraction stemming from the ionized upper part of the atmosphere (extending up to about 1500 km) and tropospheric refraction, stemming from the lower, neutral layers of the atmosphere. Ionospheric refraction is wavelength-dependent and may be (almost completely) eliminated if coherent signals are sent through the atmosphere on different carrier wavelengths. In the VLBI technique this is achieved by observing the Quasars in different wavelengths, in the Doppler- or GPS-technique the same is achieved by using two different wavelengths for signal transmission.

For microwave techniques tropospheric refraction is the ultimate accuracy-limiting component in the error budget. As opposed to range observations in the optical band, we have to take into account the so-called *wet component* of tropospheric refraction, which is highly variable in time and space. This fact forces analysts using microwave observations to introduce station and time specific parameters (or to model the effect as a random process). It allows, on the other hand, the determination of the so-called total precipitable water vapor content above an observatory with high accuracy and high temporal resolution.

The Optical Period. For centuries optical observations were the only observation type available in astronomy. In the pre-space era a series of astrometric instruments was used for the purpose of defining a terrestrial reference frame and for monitoring Earth rotation. The photographic zenith tube and the *Danjon astrolabe* were the most advanced of these instruments. They were used by the observatories contributing first to the *ILS (International Latitude Service)*, then to the *IPMS (International Polar Motion Service)* and the *BIH (Bureau International de l'Heure)* to determine the geographic latitude of a station with a precision of about 10–40 mas (milliarcseconds) in one night. We refer to [76] for more information.

The first generation of artificial Earth satellites, like Sputnik 2 and Explorer 1, was observed with optical techniques. The balloon satellites Echo 1 and 2 and PAGEOS (PAssive GEOdetic Satellite) (which could even be seen "by naked eye") were observed by a worldwide optical tracking network. These satellites were (supposedly) spherical, consisted of layers of aluminized mylar foil. Thanks to their brightness, their tracks could easily be photographed against the star background. It was not trivial to assign time-tags to specific points of the track. Much better suited from this point of view, although more difficult to track, were smaller satellites like Geos 1 (Explorer 29) and Geos 2 (Explorer 36) equipped with flash lamps allowing for tens of thousands of high-precision optical observations. Obviously, the quasi-simultaneity of observations from different sites was easily achievable.

Fascinating results came out of this first phase of satellite geodesy. The geodetic networks on different continents could be related to the geocenter (and thus to each other) with an accuracy of about 5 meters. First reliable coefficients of the gravity field (spherical expansion up to degree and order 12–15) could be also derived.

The classical astrometric technique, i.e., the establishment of the directions from an observatory to celestial objects, was applied in the 1960s and 1970s to artificial satellites and had serious deficiencies. At that time the star catalogues were not of sufficiently good quality and the reduction time (time interval between observation and availability of results) was of the order of a few weeks in the best case. This aspect and the advent of new observation techniques promising higher accuracy ruled out the astrometric technique for a number of important applications. The optical technique no longer played a significant role in space geodesy after about 1975.

In view of newly developed observation techniques, e.g., CCD, and much better star catalogues based on astrometry missions (e.g., the HIPPARCOS (HIgh Precision PARallax COllecting Satellite) mission, named in honour of the ancient Greek astronomer Hipparchus (ca. 190–125 B.C.)), the optical observations are likely to play a more prominent role in space geodesy in the near future. Additional evidence supporting this statement will be provided in Chapter 8.

The Doppler Period. The *NNSS (U.S. Navy Navigation Satellite System)*, also called *TRANSIT* system (after the survey transit instrument), had a significant impact on the development of space geodesy. It proved that a system based on the measurement of the Doppler shift of a signal generated by a stable oscillator on board of a satellite could be used for relative positioning with remarkably high accuracies (0.1–0.5 m relative, about 1 m geocentric). The satellites sent information on two carrier frequencies (400 MHz and 150 MHz) near the microwave band.

The two frequencies allowed for a compensation of ionospheric refraction. Rather small receivers connected to omni-directional antennas made the technique well suited to establish regional and global geodetic networks. Observation periods of a few days were required to obtain the above mentioned accuracy.

The NNSS satellites were in polar, almost circular, orbits about 1100 km above the Earth's surface. Only one satellite could be observed by one receiver at a particular epoch. As opposed to astrometry the Doppler measurement technique is weather-independent. Until a significant part of the GPS was deployed (around 1990) the NNSS played a significant role in space geodesy. Many Doppler campaigns were organized to establish local, regional or global networks. With the full deployment of the GPS in the 1990s the geodetic community eventually lost interest in the Doppler system. The Transit system was shut down as a positioning system in December 1996 but continued operating as an ionospheric monitoring tool. For more information concerning the Doppler system we refer to [64].

Satellite and Lunar Laser Ranging (SLR and LLR). The *Laser (Light Amplification through Stimulated Emission of Radiation)* technique, developed in the 1950s, is capable of generating highly energetic short light pulses (of a few tens of picoseconds (ps) (1 ps $= 10^{-12}$ s)). These pulses are sent out using a conventional astronomical telescope, travel to the satellite (artificial or the Moon), are reflected by special corner cubes (comparable to the rear reflectors of bicycles) on the satellite (artificial or natural) back to the telescope, where they are detected.

The actual measurement is the travel time Δt of the Laser pulse from the telescope to the satellite and back to the telescope. Apart from refraction this light travel time, after multiplication with the speed of light c in vacuum, equals twice the distance ρ_r^s between satellite and telescope at the time the light pulse is reflected from the satellite $\rho_r^s \approx \Delta t \cdot c/2$ (the observation equations are developed in more detail in Chapter 8). Today's *SLR (Satellite Laser Ranging)* technique is used to determine the distances between observatories and satellites with an accuracy of a few millimeters and, if required, with a high repetition rate (several times per second).

SLR techniques may be used for every satellite equipped with corner cubes. Figure 2.4 shows LAGEOS (LAser GEOdetic Satellite) II, a typical SLR-dedicated satellite, which was launched in 1992. Lageos II is a spherical satellite with a diameter of 0.6 m and a weight of 405 kg. 426 corner cubes are inlaid in its surface. Lageos II is a close relative of Lageos I, which was launched in 1976. The two Lageos satellites are in stable, almost circular orbits about 6000 km above the surface of the Earth.

The two Lageos satellites are primary scientific tracking targets for the *ILRS (International Laser Ranging Service)*. The two satellites have contributed in a significant way to the determination of the Earth's gravity field. Many more targets are regularly observed by the ILRS. Some of them, like the French low orbiting satellite Starlette, with a diameter of 24 cm, are similar in design to the Lageos satellites and serve a similar purpose. For other satellites the SLR technique is just the primary or backup technique for precise orbit determination.

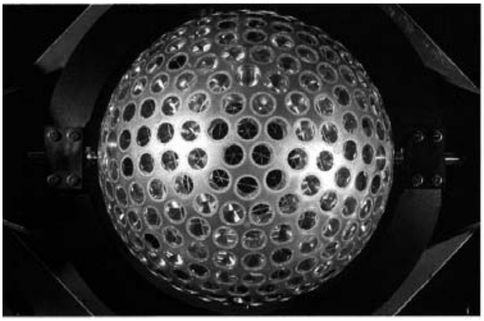


Fig. 2.4. The Lageos 2 spacecraft

With the exception of UT1 (Universal Time, defined by Earth rotation), the SLR technique is capable of determining all parameters of geodetic interest (station coordinates and motion, Earth rotation parameters, gravity field). The unique and most valuable contributions lie in the determination of the Earth's (variable) gravity field, in the determination of the geocenter (i.e., the location of the polyhedron formed by connecting the SLR stations with respect to the geocenter), and in calibrating geodetic microwave techniques.

From the technical point of view there is no principal difference between SLR and *LLR (Lunar Laser Ranging)*: Light travel times are measured from the observatory to one of the laser reflectors deployed on the Moon by the Apollo space missions or the Russian unmanned Lunokhod missions. The scientific impact of LLR is significant. LLR was, e.g., capable of measuring directly the secular increase of the Earth-Moon distance (3.8 cm per year), an effect which is in turn coupled with the deceleration of the angular velocity of Earth rotation. Also, LLR is well suited to evaluate gravitational theories.

Very Long Baseline Interferometry. Today, VLBI is the only non-satellite geodetic technique contributing data to the IERS. The principles of the technique are briefly outlined in Chapter II- 2.

Its unique and fundamental contribution to geodesy and astronomy consists of the realization of the inertial reference system and in the maintenance of the long-term stability of the transformation between the celestial and terrestrial reference frame.

The ICRS (International Celestial Reference System) and the ICRF (International Celestial Reference Frame), the realization of the ICRS, are defined and maintained by the IERS [4]. It was adopted by the IAU as the primary celestial reference system replacing the optical predecessors.

An accurate and stable celestial reference frame is a prerequisite for the establishment of a terrestrial reference frame. In this sense VLBI plays a decisive role for the definition and maintenance of the terrestrial reference frame, as well, and for establishing the transformation between the two frames. VLBI is the only technique providing the difference UT1–UTC (where UTC (Universal Time, derived from atomic clocks)) with state-of-the-art accuracy and excellent long-term stability. Also, VLBI is the only technique capable of determining precession and nutation, defining the position of the Earth's rotation axis in the inertial system, with an angular resolution below the mas-level.

The observation and analysis aspects are today coordinated by the IVS (International VLBI Service for Astrometry and Geodesy) (see Table 2.5).

The Global Positioning System (GPS). Today, the GPS is the best known of the space geodetic techniques. The system has a deep impact on science *and* on society reaching far beyond space geodesy. GPS revolutionized surveying, timing, car and aircraft navigation. Millions of hand-held receivers are in use for navigation. Spaceborne applications of the GPS are about to revolutionize geodesy and atmosphere sciences.

GPS is a navigation system allowing for instantaneous, real-time, *absolute* positioning on or near the surface of the Earth with an accuracy of few meters. An unlimited number of users may use the system simultaneously. *Absolute* means that the estimated coordinates may be established using only one receiver and that they refer to a geocentric Earth-fixed coordinate system. This coordinate system, the WGS-84 (World Geodetic System 1984), is today aligned with sub-meter accuracy to the ITRF (International Terrestrial

Reference Frame). The ITRF is the realization of the ITRS (International Terrestrial Reference System), is maintained by the IERS.

The space segment of the GPS nominally consists of 24 satellites (21 operational satellites plus 3 active spares). The satellites are in almost circular orbits distributed in six planes approximately 20000 km above the Earth's surface. These planes are separated by 60° on the equator and inclined by 55° with respect to the equator. The constellation is illustrated in Figure 2.5 as it would be seen from outside the system from the poles (left) and from a latitude of 35°.

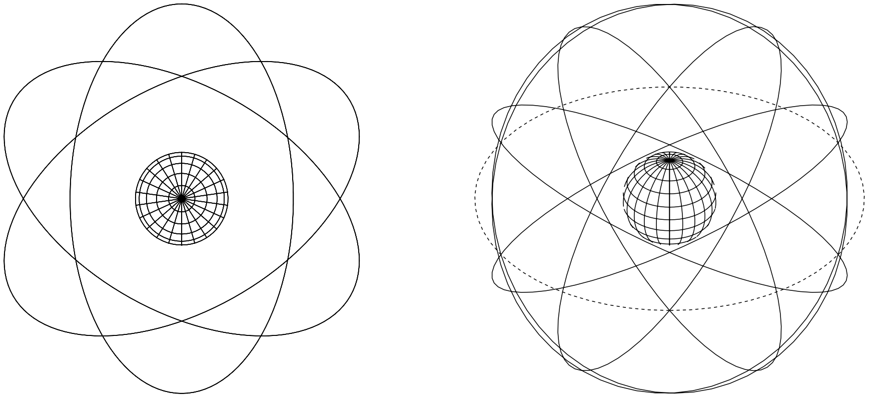


Fig. 2.5. The GPS as seen from geographic latitudes of 90° and 35°

The revolution period is half a sidereal day ($11^{\mathrm{h}}58^{\mathrm{m}}$), which means that for a given location on the Earth's surface the satellite constellation above horizon repeats itself after one sidereal day (solar day minus four minutes). Figure 2.6 shows a Block II satellite. The first full GPS generation was realized around the mid 1990s with this type of GPS satellites. We distinguish the main body of the satellite with the antenna array pointing to the center of the Earth and the solar panels. The attitude is maintained by *momentum wheels*, which have to guarantee that the antenna array is always pointing to the center of the Earth and that the solar panel axes are perpendicular to the Sun-satellite direction. The satellite is capable of rotating the solar panels into a position perpendicular to the same direction.

The GPS satellites transmit (essentially) the same information on two coherent carrier frequencies $L1$ and $L2$ (with wavelengths of about 19 and 24 cm). The two carriers are used to model (or eliminate) the frequency-dependent part of the signal delay caused by the Earth's upper atmosphere, the ionosphere. Both carriers are coherent, i.e., they are generated by one and the same highly stable oscillator onboard the satellite. The information is trans-

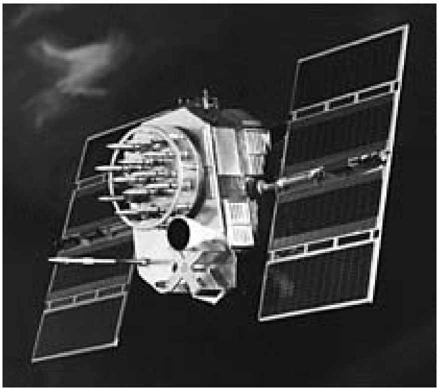


Fig. 2.6. Block II GPS satellite

mitted by the phase modulation technique. Two observables, the so-called *code observable* and the *phase observable* are used for GPS positioning and navigation. The code observable corresponds to the distance between the satellite position (at signal transmission time) and the receiver position (at signal reception time). The code observable is biased by the clock errors of the satellite and the receiver, and atmospheric transmission effects. The GPS code available for the civilian user community is accurate to about 1–3 m. The phase observable is based in principle on a count by the receiver of the incoming carrier waves (integer number plus fractional part). The measured quantity is closely related to that of the code observable: exactly as the code observation it corresponds to the distance between satellite and receiver, but it contains one more bias, an initial phase count (the receivers have to start their counts with an arbitrary value). The trouble introduced by the additional unknown is counterbalanced by the extremely high accuracy of the GPS phase observable: the phase observable is established with *mm*- rather than *m*-accuracy.

The phase observations allow it to establish local GPS networks with *mm*-accuracy, regional and global networks with about *cm*-accuracy. This is only possible, if precise satellite orbit and clock information, such as generated and distributed by the *IGS (International GPS Service)*, is available. Figure 2.7 shows the IGS network as of January 2002.

Over 200 IGS sites, distributed all over the globe, permanently observe all satellites in view and transmit their observations (at least) on a daily basis to the IGS Data Centers.

The data are then analyzed by IGS Analysis Centers, which deliver *rapid* and *final* products. Rapid IGS products are available with a delay of about one day (or even below), final products with a delay of about ten days. Daily products include satellite orbits with an accuracy of about 5 cm, satellite clocks with an accuracy of about 0.1 ns (nanoseconds), daily values of polar motion components accurate to about 0.1 mas, (corresponding to 3 mm on the Earth's surface), and LOD (Length of Day) estimates with an accuracy of about 20 μs/day (microseconds per day). These products are essential contributions to the monitoring of Earth rotation.

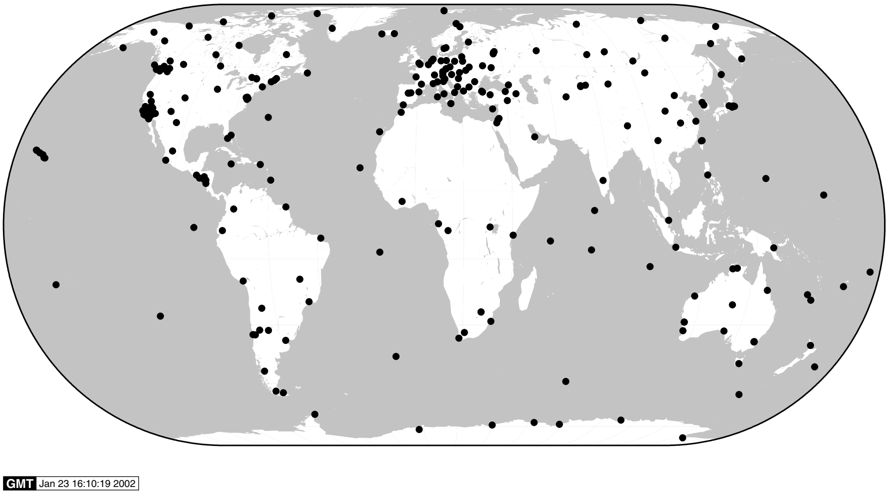


Fig. 2.7. The International GPS Service (IGS) Network in 2002

In addition the IGS Analysis Centers perform weekly global coordinate solutions of their portion of the IGS network. These results are used, together with the results of the other space techniques, for the establishment of the ITRF.

The IGS products (orbits, satellite and receiver clock corrections, Earth rotation parameters, coordinates and velocities of IGS stations) are used as known *a priori* information to establish regional networks for crustal deformation studies. More and more, the IGS network is used for purposes other than space geodesy. Let us mention that the IGS network has been enhanced to include time and frequency transfer and that it is able to monitor the ionosphere.

From the point of view of space geodesy GPS is a "work horse" with important contributions to the establishment and maintenance of a dense terrestrial reference frame, providing Earth rotation parameters with a high time resolution. It should not be forgotten, that the GPS – like every satellite geodetic method – is not able to maintain a long-term stability of UT1 or of precession and nutation. Moreover, despite the fact that GPS is a satellite geodetic technique, it is not well suited to determine the Earth's gravity field. The height of the GPS satellites is one of the limiting factors.

The GPS observables and some of the essential results derived from them are presented in section 8.5.3. Specific examples, e.g., concerning Earth rotation, will be included in other chapters of this book. For detailed information concerning the GPS as a tool for geodesy and geodynamics the readers are referred to [122]. The interdisciplinary aspects offered by the GPS in general and the IGS in particular are also discussed by Beutler et al. in [17].

Other Satellite Microwave Techniques. The Russian GLONASS (GLObal NAvigation Satellite System) is so closely related to the GPS that there is a number of combined GPS and GLONASS receivers available. These receivers were used in the first global GLONASS tracking and analysis campaign, the *IGEX-98* (International GLONASS Experiment 1998). The experiment revealed that a combined analysis of GPS and GLONASS is promising for science and navigation.

The French DORIS (Doppler Orbitography by Radiopositioning Integrated on Satellite) proved to be a very powerful tool for orbit determination. It is, e.g., one of the orbit determination systems used in the TOPEX/Poseidon (TOPEX (TOPography EXperiment for Ocean Circulation)) mission (see below). Also, DORIS possesses a very well designed ground tracking network. This is one reason why DORIS was accepted as an official space technique by the IERS (see Table 2.5).

The German PRARE (Precise Range And Range-rate Equipment) system may be viewed as the German counterpart of the DORIS system. It is used as an orbit determination tool, e.g., on the European Space Agency's ERS-2 (Earth Remote Sensing 2) spacecraft.

The Galileo system, to be implemented by the ESA (European Space Agency) in the first decade of the 21st century, will soon be added to the list of powerful operational satellite navigation systems.

Geodetic Satellite Missions. There were many satellite missions in the past and there will be more in the future in which the satellite is used as an observing platform to study aspects of the Earth relevant to geodesy and geodynamics. Let us mention in particular that *altimetry missions* significantly improved our knowledge of the sea surface topography, ocean currents, tidal motions of the oceans, etc.

Figure 2.8 shows the TOPEX/Poseidon spacecraft. The mission is a combined U.S. and French altimetry mission. It is the first mission which was specially designed to investigate ocean currents. One entire volume of the

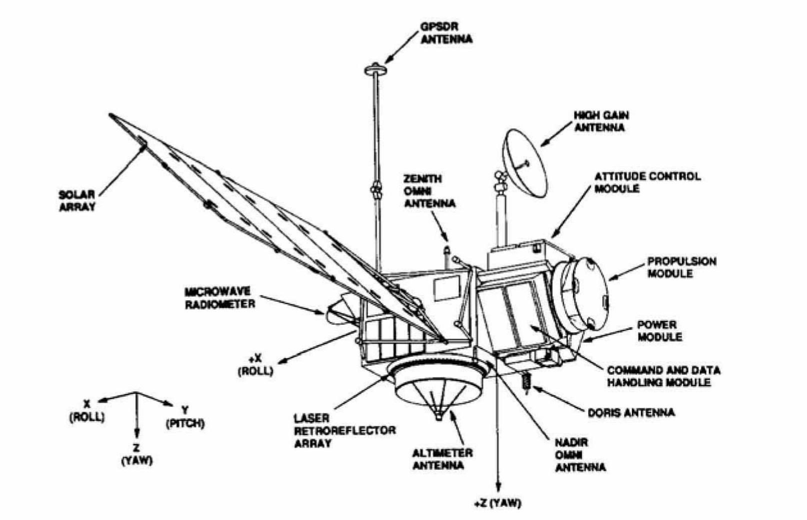


Fig. 2.8. The TOPEX/Poseidon spacecraft

Journal of Geophysical Research was devoted to this mission [123]. For space geodesy the TOPEX/Poseidon mission was a kind of "rosetta stone mission" because its orbit was determined using three independent systems, the French DORIS system, SLR tracking, and the GPS. All three systems proved their capability. The radial component of the orbit (which is of crucial importance for altimetry missions) could be established with an accuracy of a few cm. JASON is the TOPEX/Poseidon follow-on mission. It is in orbit since early 2001. JASON, exactly as TOPEX/Poseidon, is a NASA(JPL)/CNES mission, named after the mythological hero who led the Argonauts on the search for the Golden Fleece. According to information available through the ILRS (internet address in Table 2.5) JASON symbolizes both the hard-fought quest for a worthy goal and civilization's fascination with the ocean and its mysteries. The specification of "1" attests to the expectation that JASON is one of a series of TOPEX/Poseidon follow-on missions.

For geodesy, geodynamics, and atmosphere physics the CHAMP (CHAllenging Mini-satellite Payload for geophysical research and application, German mission), GRACE (Gravity Recovery And Climate Experiment, U.S. / German mission), and the upcoming GOCE (Gravity field and Ocean Current Explorer, ESA mission) are and will be fundamental. It is expected that our knowledge of the gravity field (using spaceborne GPS receivers, accelerometers, gradiometers) to measure the non-gravitational forces and gravity gradients will significantly increase thanks to these missions.

Also, CHAMP and GRACE are able to produce atmosphere profiles using the *GPS occultation method*: the signal (phase and code) of a GPS satellite is monitored by a spaceborne GPS receiver on a LEO (Low Earth Orbiter) during the time period the line of sight LEO → GPS satellite scans through the Earth's atmosphere. These developments support our initial statement that interdisciplinary aspects are becoming more and more important in Space Geodesy.

Organizations. Table 2.5 gives an overview of the international scientific organizations, which are relevant for the worldwide coordination in space geodesy.

They all are *IAG (International Association of Geodesy)* services. The IERS and the IVS are in addition IAU services. The IERS, IGS and IVS are members of FAGS (Federation of Astronomical and Geodesical Data Analysis Services).

IGS, ILRS and IVS are technique-specific services. The IERS is a multi-technique service. It was established in 1988 as the successor of the International Polar Motion Service (IPMS) and the Earth rotation branch of the Bureau International de l'Heure (BIH). The IERS products are based on the measurements and products of the technique-specific services.

CSTG (Commission on Coordination of Space Techniques) is a commission of IAG and a subcommission of COSPAR (Committee on Space Research), the Commission on Space Research. It has a coordinating function within space geodesy. In the time period 1995–1999 the CSTG created the ILRS and the IVS, and it organized the first global GLONASS experiment IGEX-98 together with the IGS.

More information about these services may be found at the internet addresses in Table 2.5.

Protagonists of Space Geodesy. It would be easily possible to create a list of eminent pioneers of this field (comparable to Table 2.1). Because a short list, written by someone deeply involved in space geodetic research for more than thirty years, would most likely be highly subjective, no such table is provided here.

Table 2.5. Space-geodetic services

Acronym	Name, Mission, Internet
CSTG	Commission on International Coordination of Space Techniques. Coordination between space geodetic organizations, organizes projects.
IERS	International Earth Rotation Service. Establishes and maintains celestial and terrestrial reference frame, generates combined Earth orientation parameter series http://hpiers.obspm.fr
IGS	International GPS Service. Makes available GPS data from its global network, producing and disseminating high accuracy GPS orbits, Earth rotation parameters, station coordinates, atmospheric information, etc. http://igscb.jpl.nasa.gov
ILRS	International Laser Ranging Service. Collects, archives, and distributes SLR and LLR datasets. Generates scientific and operational products http://ilrs.gsfc.nasa.gov
IVS	International VLBI Service for Geodesy and Astrometry. Operates and supports VLBI programs. Organizes geodetic, astrometric, geophysical research and operational activities. http://ivscc.gsfc.nasa.gov

3. The Equations of Motion

Basic concepts related to space, time, matter, and gravitation are briefly addressed in Section 3.1. In order to simplify the discussion we will first assume the celestial bodies to be either point masses (planetary system) or rigid bodies (satellite geodesy and Earth rotation). The assumption of rigidity will be dropped in sections 3.3.7 and 3.3.8, where the basic properties of a non-rigid Earth and its rotation are introduced. The structure of the equations of motion does not change too much. Some physical entities (like the inertia tensor and the angular momentum of a planet) have to be modified to take changes due to deformations into account. The mathematical structure, however, remains pretty much the same for a rigid and a deformable celestial body – at least as long as the deformations are small.

The equations of motion for a system of point masses are developed in section 3.2. Only the gravitational forces according to the inverse square law (2.1) are taken into account. It is assumed that the mass of one of the bodies dominates those of all the others, i.e., the resulting equations of motion refer to a planetary system. From these equations the well-known ten first integrals are derived.

In Section 3.3 the equations of motion for the system Earth-Moon-Sun are set up under the assumption that Earth and Moon are rigid celestial bodies of finite dimensions with given density distributions. The equations of motion for the orbital and for the rotational motions of these bodies are developed directly from the Newtonian axioms (Table 2.2) and the law of universal gravitation (2.1). The equations for the orbital and the rotational motions are coupled. Because the coupling mechanism is only weak, it is possible to derive handy approximations for the orbital and the rotational motion of the system Earth-Moon-Sun.

In Section 3.4 the equations of motion for an artificial Earth satellite are presented. For LEOs many terms of the Earth's gravity field have to be taken into account. Also, as opposed to the other equations of motion studied, non-gravitational forces play a key role. For these reasons, the equations of motion associated with an artificial Earth satellite are the most complex considered here.

We are mainly concerned with classical mechanics in this book. One should keep in mind, however, that the equations of motion should be studied in the framework of the theory of general relativity. The result of such studies would be rather complex. Fortunately it is possible to account for the relativistic effects approximately by "slightly modifying" the Newton-Euler equations of motion. Relativistic versions for the equations of motion for the planetary system and for artificial satellites are introduced in section 3.5.

The equations of motion derived in this Chapter are reviewed in Section 3.6. The mathematical structure of the equations of different types is compared, common aspects and differences are discussed.

3.1 Basic Concepts

The term *inertial reference system* or simply *inertial system* is defined as a time scale in the Newtonian sense and a rectangular Cartesian coordinate system (named in honour of René Descartes (1596–1650)) in the three-dimensional Euclidean space $\mathbb{E}^3$, in which celestial bodies obey Newton's laws of motion (Table 2.2) and his law of universal gravitation (2.1).

Time is the independent argument in the equations of motion. Newton used the term "absolute time" to distinguish it from measures of time which are far from uniform, like e.g., solar time as given by a Sun dial. Every "strictly" periodic phenomenon may be used to realize an absolute time scale – by counting the number of periods elapsed since a conventional time origin. As one wishes to use the best possible periodic phenomenon for the purpose, it is usually impossible to decide whether the underlying phenomenon is strictly periodic or not. Such a decision can only be made as soon as a "better clock" becomes available.

UT (Universal Time) is a measure of time reflecting the mean diurnal motion of the Sun. Formally, UT is derived from ST (Sidereal Time), which is in turn a measure of time defined by the apparent diurnal motion of the stars. Both, ST and UT, are time scales based on the Earth's rotation. UT (and ST) are determined from observation. Due to the effect of polar motion (see Chapter II- 2) UT, as observed at a particular observatory on the Earth, slightly depends on the location of the observatory. This realization of UT is designated as UT0. UT1 (UT corrected for polar motion effects) is in essence the best possible "absolute time", which can be obtained from Earth rotation.

Up to 1960 UT was the official realization of absolute time in astronomy. From 1960 to 1984 ET (Ephemeris Time), based on the orbital motion of the Moon and the planets, was used for the same purpose. ET was determined from the equations of motion for the bodies in the planetary system. It is by definition the best possible time scale for the purpose of Celestial Mechanics.

Unfortunately, ET was only available months after real time. This circumstance and the fact that since about 1950 atomic clocks started to provide a time scale of highest accuracy and stability, led in 1984 to the introduction of atomic time scales in astronomy.

The atomic time scale is called TAI (International Atomic Time). TAI is realized by an ensemble of atomic clocks distributed worldwide at national and international timekeeping laboratories. These clocks are able to reproduce the second (s) of the SI (International System of units). The clock combination used to define TAI asks for corrections due to special and general relativity. TAI corresponds to the performance of a perfect clock situated on the geoid, the equipotential surface of the Earth at sea level. TT (Terrestrial Time), which is today used as independent argument for Celestial Mechanics in the Earth-near space is derived from TAI by the following equation:

$$\mathrm{TT} = \mathrm{TAI} - 32.184 \text{ s} . \tag{3.1}$$

TT is a time scale uniquely based on atomic time. TT is measured in units of days defined as 86400 s in the SI. TAI and TT serve as uniform time-scales in the non-relativistic approximation of the equations of motion referring to Celestial Mechanics problems in satellite geodesy. The constant -32.184 was introduced to make the transition from ET to TAI smooth at one particular epoch in time, namely January 1, 1958.

UTC (Universal Time Coordinated) agrees with UT to within one second. UTC is derived from TAI by adding an integer number $N(\mathrm{TAI})$ of *leap seconds* to TAI to guarantee that $\mid \mathrm{UT1} - \mathrm{UTC} \mid < 1$ s . Formally, one may write

$$\mathrm{UTC} = \mathrm{TAI} + N(\mathrm{TAI}) . \tag{3.2}$$

The leap seconds are introduced (if required) end of July and/or end of December. Leap seconds and $N(\mathrm{TAI})$ are announced by the IERS in its Bulletin C.

Several other time scales are in use in satellite geodesy, GPS time probably being the best known. GPS time and UTC (thus also between GPS time and TAI) differ by an integer number of seconds.

TDB (Barycentric Dynamical Time) is the independent argument of the equations of motion referring to the barycenter of the solar system. The precise definition of TDB depends on the gravitational theory used. For most applications it is sufficient to use the approximation (see [107]):

$$\mathrm{TDB} = \mathrm{TT} + 0.001658 \;\sin g + 0.000014 \;\sin 2g \ , \tag{3.3}$$

where the amplitudes are given in seconds and

$$g = 357.53^\circ + 0.9856003^\circ \ (\mathrm{JD} - 2451545.0) \ , \tag{3.4}$$

where JD (Julian Date).

For more information concerning the time scales actually used in astronomy we refer to [107]. A concise overview of different time scales in use today may also be found in [75].

For future reference we include the numerical values for the gravitational constant in Newton's law of universal gravitation (2.1). In the SI, where meters (m), kilograms (kg), and seconds (s) are used as units for length, mass, and time, the constant of gravitation has the following value:

$$G = 6.67259 \times 10^{11} \ \mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2} \ . \tag{3.5}$$

In applications related to the planetary system it is tradition to use other units for time, mass, and length: One day d of 86400 s (SI) (in essence one solar day) is used as time unit, the mass of the Sun ($m_\odot = 1.9891 \times 10^{30}$ kg) as mass unit; the mean distance between the center of mass Earth-Moon and the Sun, originally defined to be the AU (Astronomical Unit), serves as length unit. In these units the constant of gravitation is written as k^2, where k is referred to as the *Gaussian constant.* Today, the AU is defined in such a way that the Gaussian constant k keeps the same numerical value, namely

$$k = 0.01720209895 \sqrt{(\mathrm{AU})^3 m_\odot^{-1} d^{-2}} \ , \tag{3.6}$$

as it already had at Gauss's epoch. This is why every astronomer knows the numerical value of k by heart. It is handy to assign a constant value to the product "solar mass · gravitation constant". The drawback lies in the fact that the semi-major axis a of the center of mass of the Earth-Moon system in its orbit around the Sun can no longer be strictly interpreted as $a = 1$ AU. Currently, the semi-major axis a of the Earth-Moon barycenter is considered to have a length of (see [107]):

$$a = 1.00000003 \ \mathrm{AU} \ . \tag{3.7}$$

The constant k is also approximately the mean daily motion of the center of mass of the Earth-Moon system in its orbit around the Sun, expressed in rad/d, what corresponds to $180°/\pi \cdot k \approx 0.986$ °/d .

Newton's absolute space corresponds to what we call today *inertial space.* In mathematical terms this space is the three-dimensional Euclidean space $\mathbb{E}^3$, where we may introduce a rectangular, right-handed Cartesian coordinate system. Such a coordinate system is defined by three orthogonal unit vectors $\boldsymbol{e}_i, \ i = 1, 2, 3$, originating from an origin O. For a right-hand system we have

$$\boldsymbol{e}_3 = \boldsymbol{e}_1 \times \boldsymbol{e}_2 \ ,$$

where $\boldsymbol{e}_1 \times \boldsymbol{e}_2$ is the vector product of the two vectors $\boldsymbol{e}_i, \ i = 1, 2$. The set of the three unit vectors $\boldsymbol{e}_i, \ i = 1, 2, 3$, is also called an orthonormal base in $\mathbb{E}^3$. A vector $\boldsymbol{x}$ pointing from the origin of the coordinate system to an arbitrary point in $\mathbb{E}^3$ is called a position vector. Each position vector may be represented as a linear combination of the three base vectors:

$$\boldsymbol{x} = x_1 \; \boldsymbol{e}_1 + x_2 \; \boldsymbol{e}_2 + x_3 \; \boldsymbol{e}_3 \ . \tag{3.8}$$

The coefficients $x_i, \ i = 1, 2, 3$, are the Cartesian coordinates of the vector $\boldsymbol{x}$ in the coordinate system defined by the unit vectors $\boldsymbol{e}_i, \ i = 1, 2, 3$. It should be noted that the vector $\boldsymbol{x}$ is a quantity independent of the specific coordinate system, whereas the coordinates $x_i, \ i = 1, 2, 3$, are related to the specific coordinate system. This is documented by the following relation emerging from eqn. (3.8) by scalar multiplication with the unit vector $\boldsymbol{e}_i$:

$$x_i = \boldsymbol{e}_i \cdot \boldsymbol{x} \ , \quad i = 1, 2, 3 \ . \tag{3.9}$$

It is convenient to use the following matrix notation for the coordinates of $\boldsymbol{x}$

$$\boldsymbol{x}_e^T \overset{\text{def}}{=} (x_1, x_2, x_3) \ . \tag{3.10}$$

$\boldsymbol{x}_e$ is a column-matrix with three elements, $\boldsymbol{x}_e^T$ is its transpose, a row matrix with three elements. We will also call $\boldsymbol{x}_e$ the coordinate matrix.

We should make a clear distinction between a vector and its coordinate matrix. In an attempt to reduce the formalism to the absolute minimum, we will often leave out the index specifying the coordinate system (above, we used e as an example), if no misunderstandings are possible.

Let us assume that $\boldsymbol{x}$ is the position vector of a spherically symmetric celestial body (in the sense of Theorem 8 in Table 2.2), i.e., we assume that the mass distribution in the body is spherically symmetric w.r.t. the center of mass of the body, or that the size of the body is very small compared to the distances between the bodies. In the former case we interpret $\boldsymbol{x}$ as the position vector of (the center of mass of) the body, in the latter case we speak of a point mass. Our goal is the derivation of the *trajectories* (see Figure 3.1) of all bodies (or point masses) of a mechanical system (rigid, spherically symmetric bodies, or point masses) as a function of time.

The realization of a coordinate system is a *coordinate frame.* How is the inertial coordinate system realized? Taking the same pragmatic standpoint as in the case of the realization of the uniform timescale we postulate that the frame does not rotate with respect to the ensemble of Quasars. This realization actually is closely related to definition 2) (out of four) of an inertial frame in [127]. We demand that the unit vectors $\boldsymbol{e}_i, \ i = 1, 2, 3$, do not rotate w.r.t. this ensemble of Quasars. This does not yet imply, however, that the system is inertial, because linear accelerations, e.g., along one axis, still might occur. In order to exclude such motions we have to set up the equations of motion and derive the origin of the system as a function of time. This will be a byproduct of the next section. An inertial reference system realized in this way is called an *inertial reference frame.*

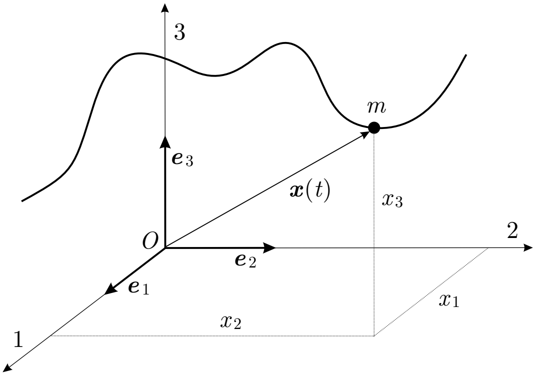


Fig. 3.1. Trajectory of a point mass m in the inertial system

When establishing the inertial reference frame we neither need the equatorial plane of the Earth nor the ecliptic, the plane of the orbital motion of the center of mass of the Earth-Moon system around the Sun. The ecliptic does, however, play an important role in Celestial Mechanics of the planetary system, and the Earth's equatorial plane plays a similarly important role in studies of the Earth-near space. The role of the former plane is given by the circumstance that the inclinations of the orbital planes of most planets w.r.t. the ecliptic are small (see Table II- 4.1), the role of the latter is given by the flattening of the Earth with the equatorial plane as symmetry plane. It makes sense to define the unit vectors $\boldsymbol{e}_i,\ i = 1, 2, 3$, in such a way that the first unit vector $\boldsymbol{e}_i$ lies in the intersection of the equatorial and the ecliptic plane. This axis points into the direction of the vernal equinox. For studies in the planetary system $\boldsymbol{e}_2$ is best defined to lie in the ecliptic, 90° away from the vernal equinox in the direction of the motion of the Earth, for Celestial Mechanics in the Earth-near space $\boldsymbol{e}_2$ is defined to lie in the equator, 90° away from the vernal equinox in the direction of the rotation of the Earth.

Both, the equatorial plane (due to precession and nutation) and the ecliptic (due to planetary perturbations) are rotating w.r.t. the inertial system. This is why we have to specify an epoch when defining the coordinate frames using the equator and ecliptic. Today, the epoch $J2000.0$ serves as normal epoch. $J2000.0$ corresponds to January 1, 2000, 12^{h} UTC (for more details we refer to [107]).

3.2 The Planetary System

We assume that N point masses are moving uniquely under their gravitational attraction and that there are no other celestial bodies (masses) outside the system. These assumptions define the classical N-body problem.

Let us furthermore assume that $N = n + 1$. Each point mass m_i is fully described by its mass m_i and its position vector $\boldsymbol{x}_i(t)$, $i = 0, 1, \ldots, n$, for all times t. Let us furthermore assume that the mass m_0 dominates the others. This is the case in our planetary system where m_0 corresponds to the Sun.

If we follow the trajectory of one of the bodies as a function of time we may also define the *velocity* $\dot{\boldsymbol{x}}_i$ of each point mass as the first time derivative of the position vector:

$$\dot{\boldsymbol{x}}_i = \frac{d\boldsymbol{x}_i(t)}{dt} \overset{\text{def}}{=} \lim_{\Delta t \to 0} \frac{\boldsymbol{x}_i(t + \Delta t) - \boldsymbol{x}_i(t)}{\Delta t} \ , \quad i = 0, 1, 2, \ldots, n \ . \tag{3.11}$$

If the position vectors $\boldsymbol{x}_i(t)$ and the velocity vectors $\dot{\boldsymbol{x}}_i(t)$ at the initial epoch $t = t_0$ are known, we have the task of finding the trajectories $\boldsymbol{x}_i(t)$ for all time arguments t and for all point masses m_i, $i = 0, 1, 2, \ldots, n$, in the inertial system. The ensemble of vectors $\boldsymbol{x}_i(t_0)$, $\dot{\boldsymbol{x}}_i(t_0)$, $i = 0, 1, \ldots, n$, represents the initial state vector of the entire system.

3.2.1 Equations of Motion of the Planetary System

According to (the modern understanding of) Newton's second law, his corollary concerning the superposition of forces (see Table 2.2), and the law of universal gravitation (2.1) (replacing, however, for our application the gravitation constant G by k^2, see eqn. (3.6)) we may write down the equations of motion in the inertial system for the $N = n + 1$ point masses:

$$\frac{d\,(m_i \dot{\boldsymbol{x}}_i)}{dt} = -\,k^2\, m_i \sum_{j=0, j \neq i}^{n} m_j \, \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \ , \quad i = 0, 1, 2, \ldots, n \ . \tag{3.12}$$

On the left hand side we have the first derivative of the linear momentum, on the right hand side the superposition of gravitational forces acting on point mass m_i. Assuming that the masses do not change with time (which will never be 100% true) we may write

$$\ddot{\boldsymbol{x}}_i = -\,k^2 \sum_{j=0, j \neq i}^{n} m_j \, \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \ , \quad i = 0, 1, 2, \ldots, n \ . \tag{3.13}$$

It is important to note that the mass m_i does not show up in the equation for this particular point mass. From the mathematical point of view the equations of motion (3.13) of our N-body problem form an ordinary, coupled, nonlinear differential equation system of second order in time. Mathematical analysis tells that unique trajectories exist (under certain conditions) for all times $t \in (-\infty, +\infty)$, provided the initial state of the system is known.

$$\boldsymbol{x}_i(t_0) \overset{\text{def}}{=} \boldsymbol{x}_{i0} \ , \quad \dot{\boldsymbol{x}}_i(t_0) \overset{\text{def}}{=} \dot{\boldsymbol{x}}_{i0} \ , \quad i = 0, 1, 2, \ldots, n \ . \tag{3.14}$$

The equations (3.13) are vector equations. If we want to solve them (numerically or otherwise) we have to select a concrete coordinate system with a suitable origin. There is an entire class of equivalent inertial reference frames, because the eqns. (3.13) are invariant under a *Galilei transformation* (named in honour of Galileo Galilei (1564–1642)), composed of a translation $\boldsymbol{X}_0$ and a velocity $\boldsymbol{V}_0$:

$$\begin{aligned} \boldsymbol{x}_i &= \boldsymbol{x}_i' + \boldsymbol{X}_0 + \boldsymbol{V}_0\, t \\ \dot{\boldsymbol{x}}_i &= \dot{\boldsymbol{x}}_i' + \boldsymbol{V}_0 \\ t &= t' \ . \end{aligned} \tag{3.15}$$

When introducing these transformation equations into eqns. (3.13) we easily verify that the transformed equations have an identical structure in the new reference frame, which is therefore inertial, as well.

Assuming that we were able to identify one particular inertial reference system, we may interpret the symbols $\boldsymbol{x}_i$ in the equations of motion as the column matrices of coordinates relative to this system. We should use an additional index, e.g., $\boldsymbol{x}_{\mathcal{I}_i}$ to identify the coordinate system. If no transformations are required, we may as well skip the index "$\mathcal{I}$". Except for this subtlety the equations of motion for the position vectors are formally identical with the corresponding equations for the coordinate matrices.

It was assumed that there are no masses outside the system of the N point masses. This is not perfectly true. Think, e.g., of the gravitational attraction the solar system experiences from our galaxy. When considering time periods of hundreds of millions of years such effects must be taken into account (the revolution period of the solar system around the galactic center is estimated to be about 250 million years). We do not account for such effects in this chapter and further explore the idea of an isolated system of point masses.

If the mass m_0 dominates all other masses it makes sense to rearrange the equations (3.13) to describe the motion of the system relative to the point mass m_0. For that purpose we introduce the notations:

$$\boldsymbol{r}_i \overset{\text{def}}{=} \boldsymbol{x}_i - \boldsymbol{x}_0 \ , \quad i = 1, 2, \ldots, n \ . \tag{3.16}$$

In our planetary system we call the vectors $\boldsymbol{r}_i$ the heliocentric position vectors. Starting from the equations of motion (3.13) in the inertial system we may easily set up the corresponding equations of motion for the heliocentric position vectors $\boldsymbol{r}_i$ by subtracting the equation for the point mass m_0 from the corresponding equation for the point mass m_i. We obtain:

$$\begin{aligned}\ddot{\boldsymbol{r}}_i &\overset{\text{def}}{=} \ddot{\boldsymbol{x}}_i - \ddot{\boldsymbol{x}}_0\\ &= -\,k^2 \sum_{j=0,j\neq i}^{n} m_j\,\frac{\boldsymbol{x}_i-\boldsymbol{x}_j}{|\boldsymbol{x}_i-\boldsymbol{x}_j|^3} \;+\; k^2\sum_{j=1}^{n} m_j\,\frac{\boldsymbol{x}_0-\boldsymbol{x}_j}{|\boldsymbol{x}_0-\boldsymbol{x}_j|^3}\\ &= -\,k^2 \sum_{j=0,j\neq i}^{n} m_j\,\frac{\boldsymbol{r}_i-\boldsymbol{r}_j}{|\boldsymbol{r}_i-\boldsymbol{r}_j|^3} \;-\; k^2\sum_{j=1}^{n} m_j\,\frac{\boldsymbol{r}_j}{|\boldsymbol{r}_j|^3} \quad .\end{aligned} \tag{3.17}$$

If we take out the first term (index $j = 0$) in the first sum on the right hand side of the above equation and term $j = i$ in the second sum and let these terms precede the two sums, we obtain the equations of motion for the heliocentric position vectors $\boldsymbol{r}_i$ in the following form:

$$\ddot{\boldsymbol{r}}_i = -\,k^2(m_0+m_i)\,\frac{\boldsymbol{r}_i}{r_i^3} - k^2\sum_{j=1,j\neq i}^{n} m_j \left\{\frac{\boldsymbol{r}_i-\boldsymbol{r}_j}{|\boldsymbol{r}_i-\boldsymbol{r}_j|^3} + \frac{\boldsymbol{r}_j}{r_j^3}\right\}\,, \quad i = 1,2,\ldots,n\,, \tag{3.18}$$

where $r_j \overset{\text{def}}{=} |\boldsymbol{r}_j|$. We easily see that the equations of motion (3.18) for the heliocentric motion again form a coupled second order differential equation system. Its dimension is $d = 3\,n$ and has been reduced by 3 when compared to the system (3.13) describing the motion in the inertial system.

If we interpret the above equations as equations for the coordinate matrices we see that the underlying heliocentric Cartesian coordinate system is always parallel to the original inertial Cartesian reference frame in the inertial space. The heliocentric coordinate system is, however, not inertial, because its origin follows the trajectory of the point mass m_0 (e.g., that of the Sun), which, according to the first of equations (3.13), shows non-vanishing accelerations w.r.t. inertial space.

It is important to note that we are able to study the development of a planetary system relative to the central mass without having defined the origin in the inertial system, using the equations of motion (3.18), provided the initial state in the heliocentric system is given by

$$\boldsymbol{r}_i(t_0) \overset{\text{def}}{=} \boldsymbol{r}_{i0}\;,\quad \dot{\boldsymbol{r}}_i(t_0) \overset{\text{def}}{=} \dot{\boldsymbol{r}}_{i0}\;,\quad i = 1,2,\ldots,n\;. \tag{3.19}$$

It was the achievement of the preceding centuries to determine the initial conditions (3.19) and the masses m_i, $i = 1,2,\ldots,n$, of (Sun) planets with higher and higher accuracy. The definition of a suitable origin in the inertial system was a secondary issue.

The structure of the equations of motion (3.18) may be further specified: the first term on the right hand side is called the *main term* of the force (per mass unit) acting on point mass m_i, the sum is called the *perturbation term.* This characterization is correct as long as there are no close encounters between the bodies of the system and if the ratios of the planetary (and satellite) masses to the mass of the central body $m_i/m_\odot \ll 1$ are small numbers. Table II-4.1 in Chapter II-4 shows that this assumption holds in our solar system, where the most massive planet, Jupiter, has a mass of only about 0.1 % of the solar mass.

We have stated above that the entire system is coupled. We have to modify this statement slightly. So far, we made the distinction between the central mass m_0 and all the other masses m_i, $i = 1, 2, \ldots, n$. We now add one more point mass to this system, assuming that its mass m

$$m \ll m_i \ , \quad i = 0, 1, 2, \ldots, n \ , \tag{3.20}$$

is negligible w.r.t. all other masses of the system. We denote the heliocentric position vector of the new point mass with $\boldsymbol{r}$. It is easy to add the equation of motion for this body to the N-body problem described by equations (3.13). The structure of the equations for the point masses with "finite masses" are unaffected by this procedure, and for the point mass m of negligible mass we obtain the following equations of motion in the heliocentric system (formally the equations of motion are obtained from eqns. (3.18) by setting $m_i = 0$ and by leaving out the index i):

$$\ddot{\boldsymbol{r}} = -\,k^2\, m_0\, \frac{\boldsymbol{r}}{r^3} \;-\; k^2 \sum_{j=1}^{n} m_j \left\{ \frac{\boldsymbol{r} - \boldsymbol{r}_j}{|\boldsymbol{r} - \boldsymbol{r}_j|^3} + \frac{\boldsymbol{r}_j}{r_j^3} \right\} \ . \tag{3.21}$$

The sum on the right hand side has to be extended only over the finite masses of the planetary system. The equations of motion (3.21) may be solved independently from the equations of motion (3.18) for the entire planetary system of the bodies with finite masses, or, in other words, we may consider the position vectors $\boldsymbol{r}_j$ as known functions of time on the right hand side of the above equations. The equations of motion (3.21) are, e.g., used to describe the trajectory of a minor planet or a comet. Note that in our solar system we could set $m_0 = 1$, which would further simplify the structure of the differential equation system.

The right hand sides of the equations (3.18) for body number i may be written as a gradient w.r.t. the coordinates of this body. The equations of motion (3.18) therefore may be written in the form

$$\ddot{\boldsymbol{r}}_i = \nabla_i \left\{ U_i + R_i \right\} \ , \quad i = 1, 2, \ldots, n \ . \tag{3.22}$$

where

$$U_i = \frac{k^2\,(m_0 + m_i)}{r_i} \tag{3.23}$$

and

$$R_i = k^2 \sum_{j=1, j\neq i}^{n} m_j \left\{ \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \frac{\boldsymbol{r}_i \boldsymbol{r}_j}{r_j^3} \right\} \ , \tag{3.24}$$

and where the gradient ∇_i is defined as

$$\nabla_i^T \overset{\text{def}}{=} \left(\frac{\partial}{\partial r_{i_1}}, \frac{\partial}{\partial r_{i_2}}, \frac{\partial}{\partial r_{i_3}} \right) , \tag{3.25}$$

where ∇_i^T is the transpose of ∇_i. U_i is the *force function* of the two-body problem (to be discussed below), R_i is called the *perturbation function* of body number i. Both, U_i and R_i are scalar functions of the $\boldsymbol{r}_j$, $j = 1, 2, \ldots, n$, but only the dependency on $\boldsymbol{r}_i$ is considered for body number i when taking the gradient of R_i. As mentioned in Chapter 2, the scalar perturbation function was introduced by Lagrange. It has the advantage for analytical developments that only one instead of three functions (corresponding to the three coordinates) has to be studied. If we inspect the perturbation function we see that the term of the sum corresponding to a particular planet is composed of a $1/r$-term which we would also expect in the equations referring to the inertial system. This term is called the *direct* perturbation term. The second term containing the scalar product of the perturbing and the perturbed body is called the *indirect* term. It is uniquely due to the transformation from the inertial to the heliocentric system.

The equations of motion for a point mass with negligible mass may be written in similar form

$$\ddot{\boldsymbol{r}} = \nabla \left\{ U + R \right\} , \tag{3.26}$$

where

$$U = \frac{k^2 \, m_0}{r} \tag{3.27}$$

and

$$R = k^2 \sum_{j=1}^{n} m_j \left\{ \frac{1}{|\boldsymbol{r} - \boldsymbol{r}_j|} - \frac{\boldsymbol{r} \; \boldsymbol{r_j}}{r_j^3} \right\} . \tag{3.28}$$

The gradient refers to the components of the position vector $\boldsymbol{r}$.

3.2.2 First Integrals

Ten scalar functions of the coordinates and velocities of the N-Body problem are known to be time-independent. We call such quantities *first integrals* or simply *integrals*. We also derive formulas for the time development of the so-called polar moment of inertia of the system (to be defined below). The result is called the *virial theorem*.

The developments of the entire section are based on the equations of motion (3.13) referring to an inertial reference frame.

Center of Mass. The center of mass of a system of point masses is defined as:

$$\boldsymbol{X}_0 \overset{\text{def}}{=} \frac{1}{M} \sum_{i=0}^{n} m_i \, \boldsymbol{x}_i \ , \tag{3.29}$$

where the total mass M of the system is given by:

$$M \overset{\text{def}}{=} \sum_{i=0}^{n} m_i \ . \tag{3.30}$$

If we multiply equation number i of the system (3.13) with the factor m_i/M and add up all resulting equations, the double sum on the right hand side adds up to zero. The differential equation for the center of mass in the inertial system therefore reads as:

$$\frac{1}{M} \sum_{i=0}^{n} m_i \, \ddot{\boldsymbol{x}}_i = \ddot{\boldsymbol{X}}_0 = \boldsymbol{0} \ . \tag{3.31}$$

This equation is solved by

$$\boldsymbol{X}_0(t) = \boldsymbol{X}_{00} + \boldsymbol{V}_{00}(t - t_0) \ , \tag{3.32}$$

where t_0 is an arbitrarily chosen origin of time, $\boldsymbol{X}_{00}$ is the position vector of the center of mass in the inertial system, $\boldsymbol{V}_{00}$ its velocity vector at the same time. As $\boldsymbol{X}_{00}$ and $\boldsymbol{V}_{00}$ both are defined by three quantities, we have actually found six first integrals by showing that the center of mass is moving according to eqn. (3.32).

We are now in a position to define an inertial reference frame with origin in the center of mass of our N bodies by asking that

$$\boldsymbol{X}_{00} = \boldsymbol{V}_{00} = \boldsymbol{0} \ . \tag{3.33}$$

This implies that our definition (3.29) for the center of mass reads as follows

$$\frac{1}{M} \sum_{i=0}^{n} m_i \, \boldsymbol{x}_i = \boldsymbol{0} \ . \tag{3.34}$$

Using the fact that the position vector $\boldsymbol{x}_i$ for point mass m_i can be written as the sum of the position vector of the Sun (the central mass) referred to the inertial system and the heliocentric position vector of point mass m_i

$$\boldsymbol{x}_i = \boldsymbol{x}_0 + \boldsymbol{r}_i \ , \quad i = 1, 2, \ldots, n \ , \tag{3.35}$$

and introducing this relation into the equation (3.34) and the corresponding equation for the velocities, we obtain the relations

$$\boldsymbol{x}_0 + \frac{1}{M}\sum_{i=1}^{n} m_i\ \boldsymbol{r}_i \overset{\text{def}}{=} \boldsymbol{x}_{\mathbf{0}} + \boldsymbol{R}_0 = \mathbf{0} \tag{3.36}$$

and

$$\dot{\boldsymbol{x}}_0 + \frac{1}{M}\sum_{i=1}^{n} m_i\ \dot{\boldsymbol{r}}_i \overset{\text{def}}{=} \dot{\boldsymbol{x}}_0 + \dot{\boldsymbol{R}}_0 = \mathbf{0}\ , \tag{3.37}$$

where $\boldsymbol{R}_0$ is the position vector of the barycenter as computed in the heliocentric system and $\dot{\boldsymbol{R}}_0$ its velocity. Obviously, we are in a position to compute the position of the Sun in the barycentric system (which is an inertial system) for any time t – provided we have solved the heliocentric equations of motion (3.13). This implies that by the transformation from the inertial to the heliocentric system we have not lost any information. We implicitly made use of the six integrals found above when reducing the number of (scalar) equations by three in the transition from the inertial to the heliocentric system.

Total Angular Momentum. The total angular momentum of a system of point masses is defined as

$$\boldsymbol{h} \overset{\text{def}}{=} \sum_{i=0}^{n} m_i\ \boldsymbol{x}_i \times \dot{\boldsymbol{x}}_i\ . \tag{3.38}$$

We multiply the equation of motion for point mass m_i in eqns. (3.13) with $m_i\ \boldsymbol{x}_i\times$, add the resulting equations and obtain:

$$\begin{aligned}
\sum_{i=0}^{n} m_i\ \boldsymbol{x}_i \times \ddot{\boldsymbol{x}}_i \qquad &= -k^2 \sum_{i=0}^{n}\ \sum_{j=1,j\neq i}^{n} m_i\ m_j\ \frac{\boldsymbol{x}_i \times (\boldsymbol{x}_i - \boldsymbol{x}_j)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \\
\frac{d}{dt}\left[\sum_{i=0}^{n} m_i\ \boldsymbol{x}_i \times \dot{\boldsymbol{x}}_i\right] &= k^2 \sum_{i=0}^{n}\ \sum_{j=0,j\neq i}^{n} m_i\ m_j\ \frac{\boldsymbol{x}_i \times \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} = \mathbf{0}\ .
\end{aligned} \tag{3.39}$$

From the last of the above set of equations we conclude that the total angular momentum of a system is conserved:

$$\sum_{i=0}^{n} m_i\ \boldsymbol{x}_i \times \dot{\boldsymbol{x}}_i \overset{\text{def}}{=} \boldsymbol{h}\ . \tag{3.40}$$

As the (constant) vector $\boldsymbol{h}$ is defined by three scalar quantities (e.g., the three components of $\boldsymbol{h}$) we have found three (first) integrals of the equations of motion of the N-body problem.

The plane perpendicular to the vector $\boldsymbol{h}$ is called the *invariable plane* or the *Laplacian plane*. Due to its definition it actually would be the natural plane of reference to describe the evolution of the planetary system (at least over long time periods). It would be much better suited than the ecliptic plane referred to a normal epoch like $J2000.0$, which becomes completely meaningless for epochs a few million years apart from $J2000.0$.

Total Energy. The total energy E of a mechanical system is defined as the sum of its kinetic energy T and potential energy $-U$, the absolute value of which is the force function U previously introduced.

$$E = T - U \ , \tag{3.41}$$

where

$$T = \tfrac{1}{2} \sum_{i=0}^{n} m_i \ \dot{\boldsymbol{x}}_i^2 \tag{3.42}$$

and

$$U = k^2 \sum_{i=0}^{n} \sum_{j=i+1}^{n} \frac{m_i \, m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \ . \tag{3.43}$$

The “force function” U of the system at a particular point in time equals the work that would have to be performed in order to completely dissolve the system, i.e., to remove each pair of bodies of the system to infinite distance. U may be written as:

$$U = k^2 \sum_{i=0}^{n} \sum_{j=i+1}^{n} \frac{m_i \, m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} = \frac{k^2}{2} \sum_{i=0}^{n} \sum_{j=0, j \neq i}^{n} \frac{m_i \, m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \ . \tag{3.44}$$

The total energy of our system of point masses thus equals

$$E = \frac{1}{2} \sum_{i=0}^{n} m_i \ \dot{\boldsymbol{x}}_i^2 \ - \ \frac{k^2}{2} \sum_{i=0}^{n} \sum_{j=0, j \neq i}^{n} \frac{m_i \, m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \ . \tag{3.45}$$

By multiplying equation i of the system of equations of motion (3.13) by $m_i \ \dot{\boldsymbol{x}}_i \cdot$ and by summing up all resulting equations the total energy is seen to be conserved:

$$\begin{aligned}
\sum_{i=0}^{n} m_i \ \dot{\boldsymbol{x}}_i \cdot \ddot{\boldsymbol{x}}_i \qquad &= - \, k^2 \sum_{i=0}^{n} \sum_{j=1, j \neq i}^{n} m_i \, m_j \ \frac{\dot{\boldsymbol{x}}_i \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \\
&= - \, k^2 \sum_{j=0}^{n} \sum_{i=0, i \neq j}^{n} m_i \, m_j \ \frac{\dot{\boldsymbol{x}}_j \cdot (\boldsymbol{x}_j - \boldsymbol{x}_i)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \\
&= - \, \tfrac{1}{2} \, k^2 \sum_{i=0}^{n} \sum_{j=1, j \neq i}^{n} m_i \, m_j \ \frac{(\dot{\boldsymbol{x}}_i - \dot{\boldsymbol{x}}_j) \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3}
\end{aligned} \tag{3.46}$$

$$\frac{1}{2} \frac{d}{dt} \left[\sum_{i=0}^{n} m_i \ \dot{\boldsymbol{x}}_i^2 \right] = \frac{1}{2} \frac{d}{dt} \left[k^2 \sum_{i=0}^{n} \sum_{j=0, j \neq i}^{n} \frac{m_i \, m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} \right] \ .$$

From the first to the second line the indices i and j were exchanged, from the second to the third the sign in the expression $(\boldsymbol{x}_j - \boldsymbol{x}_i)$ was changed and the summations over i and j were exchanged. In the third line the right hand side was calculated as the "mean value" of the expressions in the first and second line. In the last line the left and the right hand side of the equation could be written as total time derivatives, from which the conservation of total energy is obtained:

$$\tfrac{1}{2}\sum_{i=0}^{n} m_i\, \dot{\boldsymbol{x}}_i^2 \;-\; \tfrac{1}{2}k^2 \sum_{i=0}^{n}\;\sum_{j=0,j\neq i}^{n} \frac{m_i\, m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} = E\ . \tag{3.47}$$

Ten first integrals were found in this section, six related to the motion of the center of mass, three to the total angular momentum, and one to the total energy of the system. We were able to make excellent use of the first six integrals by solving the equations of motion in the heliocentric coordinate system. There is no obvious way to reduce the number of equations of motion using the other four integrals. All attempts in this direction ruin the symmetry and simplicity of the equations. The integrals may be used very well, however, to check the quality of analytical or numerical solutions of the equations of motion.

The question naturally arises whether there are other first integrals which might further reduce the complexity of the problems. There are, e.g., theorems due to Poincaré and Heinrich Bruns (1848–1919). Both are negative statements in the sense "if such and such assumptions hold, there are no other integrals". The assumptions are quite restrictive (which is why the theorems are not included here) and we agree with Moulton [77] who states: *The practical importance of the theorems by Bruns and Poincaré have often been overrated by those who have forgotten the conditions under which they have been proven to be hold true.*

Virial Theorem. The so-called *polar moment of inertia* of a system is defined as

$$I(t) \overset{\text{def}}{=} \sum_{i=0}^{n} m_i\, \boldsymbol{x}_i^2\ . \tag{3.48}$$

By construction $I(t)$ is always positive. It becomes infinite if one or several of the bodies escape from the system. On the other hand, if the orbits of all point masses would be concentric circles (with the center of mass as common center), $I(t)$ would be a constant. This condition certainly is not met in our planetary system, but it is not far from the truth either. From such considerations we see that $I(t)$ might be a good indicator to decide whether an N-body system is stable or not.

It is instructive to calculate the second time derivative of the polar moment of inertia:

$$\begin{aligned}
\dot{I} &= 2\sum_{i=0}^{n} m_i\ \boldsymbol{x}_i \cdot \dot{\boldsymbol{x}}_i \\
\ddot{I} &= 2\sum_{i=0}^{n} m_i\ \dot{\boldsymbol{x}}_i^2\ +\ 2\sum_{i=0}^{n} m_i\ \boldsymbol{x}_i \cdot \ddot{\boldsymbol{x}}_i \\
&= 4T\ -\ 2k^2 \sum_{i=0}^{n}\ \sum_{j=0,j\neq i}^{n} m_i\ m_j\ \frac{\boldsymbol{x}_i \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \\
&= 4T\ +\ 2k^2 \sum_{i=0}^{n}\ \sum_{j=0,j\neq i}^{n} m_i\ m_j\ \frac{\boldsymbol{x}_j \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \\
&= 4T\ -\ k^2 \sum_{i=0}^{n}\ \sum_{j=0,j\neq i}^{n} \frac{m_i \cdot m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} \\
&= 4T\ -\ 2U\ .
\end{aligned} \tag{3.49}$$

The second time derivative of $I(t)$ is a function of the total kinetic and potential energy (or of the force function), or the total energy (which is constant) and either the kinetic or the potential energy of the system:

$$\ddot{I} = 4T - 2U = 2E + 2T = 4E + 2U\ . \tag{3.50}$$

Equation (3.50) is a special form of the *virial theorem.* The above equation is *not* a self-contained differential equation: the right hand side of equation (3.50) may only be computed if either the potential or the kinetic energy are known as a function of time.

Because the total kinetic energy of the system must always be a positive quantity and because a positive $\ddot{I}$ inevitably leads to the destruction of the system, we have to ask for $E < 0$ as a necessary condition for a stable system. This condition is not sufficient, however.

For many applications, the virial theorem may be written in a more meaningful way, in particular if the solution is periodic or if all coordinates and velocities vary only within certain given limits. The former condition is not given in the planetary system, the latter just might be the case if the system is stable. In the cases mentioned it makes sense to compute the mean value of $\ddot{I}$ over a certain time interval Δt as

$$\bar{\ddot{I}}\left(t + \frac{\Delta t}{2}\right) = \frac{1}{\Delta t}\int_{t}^{t+\Delta t} \ddot{I}(t')\ dt' = \frac{1}{\Delta t}\left(\dot{I}(t+\Delta t) - \dot{I}(t)\right) = 4\bar{T}\ -\ 2\bar{U}\ , \tag{3.51}$$

where $\bar{T}$ is the mean value of the total kinetic energy of the system in the time interval $[t, t+\Delta t]$, $-\bar{U}$ is the mean value of the potential energy in the same time interval. For periodic solutions the quantity $\frac{1}{\Delta t}\left(\dot{I}(t+\Delta t)-\dot{I}(t)\right)$ is exactly zero, if Δt is equal to an integer number of periods. For many non-periodic systems we may assume

$$\frac{1}{\Delta t}\left(\dot{I}(t+\Delta t)-\dot{I}(t)\right) \to 0 \quad \text{for} \quad \Delta t \to \infty \ . \tag{3.52}$$

In practice Δt should be much longer than the longest period occurring in our system. If these conditions hold we obtain the virial theorem in the form

$$\bar{T} = \frac{\bar{U}}{2} \ , \tag{3.53}$$

i.e., averaged over long time intervals, twice the mean value of the kinetic energy equals the mean value of the force function. We might call this statement a *statistical conservation law.* The relation is well known in kinetic gas theory, but also in galactic dynamics or in the dynamics of star clusters. We expect that this relation holds approximately in the planetary system, as well.

3.3 The Earth-Moon-Sun-System

3.3.1 Introduction

Point masses are idealizations of real bodies, which are of finite (non vanishing) size. We might generalize the N-body problem by replacing all point masses by bodies of finite extensions. In view of the physical size of the planets and of the distances between them such a generalization would not make sense in Celestial Mechanics. There are, however, sub-systems where the bodies' size cannot be neglected. The system Earth-Moon-Sun, as visualized in Figure 3.2, is one important example. Figure 3.2 shows the Earth and the Moon as bodies of finite size and the Sun as a point mass. This latter approximation is justified because the mass distribution within the Sun shows almost perfect spherical symmetry and because the distance of the Sun w.r.t. the two other bodies is big compared to the sizes of the three bodies.

Figure 3.2 illustrates the notations used throughout this section. Two parallel coordinate systems, one inertial and one geocentric, are required. All position vectors referring to the inertial system are characterized by $\boldsymbol{x}_{\dots}$, all referring to the geocentric system by $\boldsymbol{r}_{\dots}$. The subscript $\odot$ refers to the Sun, $\earth$ to the Earth, and $\leftmoon$ to the Moon. The subscript p denotes a mass element (particle) of the Earth, $\wp$ one of the Moon. The symbol dm is reserved for a general mass element of a "general" celestial body. The vectors $\boldsymbol{x}_{\odot}$, $\boldsymbol{x}_{\earth}$, $\boldsymbol{x}_{\leftmoon}$, $\boldsymbol{x}_{p}$, $\boldsymbol{x}_{\wp}$ are the position vectors of the centers of mass of Sun, Earth, Moon, of a particular mass element of the Earth and one of the Moon in the inertial system, $\boldsymbol{r}_{\odot}$, $\boldsymbol{r}_{\leftmoon}$,

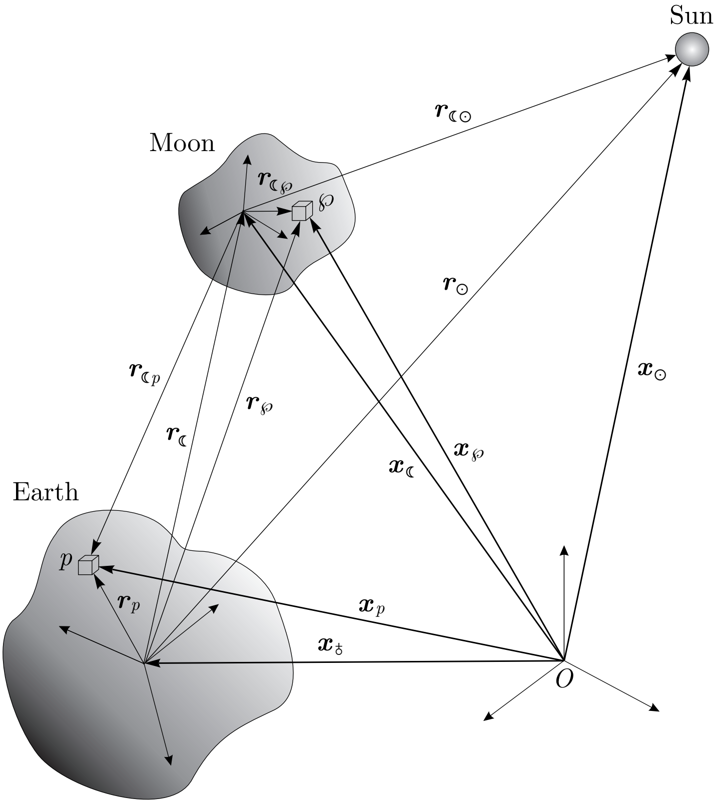

Sun
$\boldsymbol{r}_{\leftmoon\odot}$
Moon
$\boldsymbol{r}_{\leftmoon\wp}$
$\wp$
$\boldsymbol{r}_{\odot}$
$\boldsymbol{x}_{\odot}$
$\boldsymbol{r}_{\leftmoon p}$
$\boldsymbol{r}_{\wp}$
$\boldsymbol{r}_{\leftmoon}$
$\boldsymbol{x}_{\wp}$
$\boldsymbol{x}_{\leftmoon}$
Earth
p
$\boldsymbol{r}_p$
$\boldsymbol{x}_p$
$\boldsymbol{x}_{\earth}$
O

Fig. 3.2. The Earth-Moon-Sun system

$\boldsymbol{r}_p$, $\boldsymbol{r}_\wp$ are the corresponding vectors in the geocentric system. Occasionally, position vectors relative to the Moon's center of mass are required. In this case we use two symbols to denote the selenocentric position vector, e.g., $\boldsymbol{r}_{\leftmoon\odot}$ for the selenocentric position vector of the Sun, $\boldsymbol{r}_{\leftmoon\wp}$ for the selenocentric position vector of a lunar mass element.

Some of the essential facts related to the three-body problem Earth-Moon-Sun are summarized in Table II- 2.1 in Chapter II- 2. From that table one must conclude that the artist's view of the Earth-Moon-Sun system in Figure 3.2 is largely exaggerated: Earth and Moon are close to spherical, which is why the point mass model is a good approximation for the orbital motion for the (centers of mass of the) three bodies as soon as the distances between the bodies are big compared to the dimensions of the bodies.

Not only orbital, but also rotational motion has to be considered in this type of three-body problem. The basic facts (see Table II- 2.1) are well known: the Earth rotates with a period of one sidereal day about its axis. The rotation axis is inclined by about 23.5° w.r.t. the normal (pole) of the ecliptic. This inclination angle does not vary much in time (of the order of a few degrees over millions of years). It is well known that the rotation axis is precessing around the pole of the ecliptic, the period being about 26500 years. This implies that about 13000 years from now the Earth's rotation axis will point to a point 47° from today's polar star (Polaris = α Ursae Minoris), to return (more or less) to its original position after a full precession period. The regression of about $50''$ per year of the vernal equinox in the ecliptic was already known to Hipparchus.

Orbital and rotational motion of the Moon show interesting peculiarities, as well: The node of the Moon's orbital plane (its intersection with the ecliptic) regresses in the ecliptical plane with a period of about 18.6 years. It is also well established that the Moon's revolution and rotation periods agree perfectly. It is probably not common knowledge, however, that the Moon's rotation axis, inclined by about 1.54° w.r.t. the ecliptical plane also shows the effect of precession, where the precession period exactly corresponds to the period of the regression of the node, illustrating the strong correlation between orbital and rotational motion. A detailed discussion is provided in section II- 2.2.3.

It is in principle a straight forward procedure to add (some of) the planets as additional point masses to the three-body system Earth-Moon-Sun (e.g., other planets of the planetary system). This is, e.g., necessary when studying the long-term development of the Moon's orbit or of the obliquity of the ecliptic (see Chapter II- 2). Adding more point masses does not alter the mathematical structure of the problem, which is why we confine ourselves to the analysis of the three-body problem Earth-Moon-Sun subsequently.

In sections 3.3.2 to 3.3.6 it is assumed that Earth and Moon are rigid bodies. In sections 3.3.7 and 3.3.8 we will introduce the generalizations needed to discuss the rotation of a deformable Earth.

3.3.2 Kinematics of Rigid Bodies

Total Mass and Center of Mass. A body is said to be *rigid*, if the distance between any two of its mass elements remains constant in time. Assuming a continuous mass distribution described by a density function $\rho(\boldsymbol{x})$, expressed, e.g., in kg/m^3, the body's mass may be computed as a volume integral

$$m = \int dm = \int\limits_V \rho(\boldsymbol{x}_{dm})\; dV \ , \tag{3.54}$$

where the integration has to be extended over the entire volume occupied by the body.

The motion of a rigid body is completely known, if the motion of one specific mass (or volume) element of the body with position vector $\boldsymbol{x}(t)$ is known as function of time, and if the orientation of a body-fixed Cartesian coordinate system centered at this mass element is also given as a function of time (see Figure 3.3). The former motion may be called the *orbital motion*, the latter the *rotational motion.* In principle, an arbitrary point might be selected to

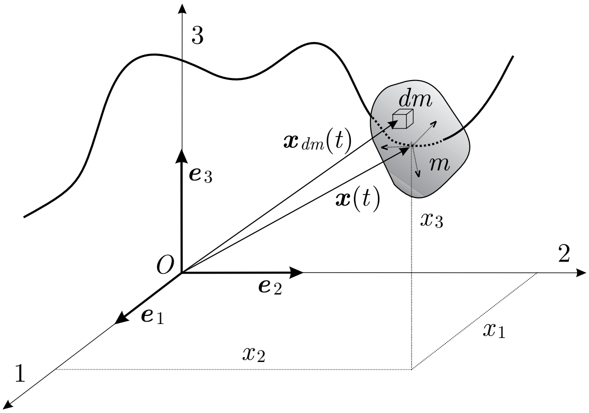


Fig. 3.3. Trajectory of an extended celestial body $\int\limits_V dm$ in the inertial system

describe the orbital motion and to serve as origin of the body-fixed coordinate system. It is convenient, however, to select the *center of mass* of the body as reference point. The center of mass of the rigid body is defined as

$$\boldsymbol{x} \overset{\text{def}}{=} \frac{1}{m} \int\limits_V \rho(\boldsymbol{x}_{dm})\ \boldsymbol{x}_{dm}\ dV\ . \tag{3.55}$$

This is a straight forward generalization of the definition for the center of mass in the N-body problem according to eqn. (3.29) – the system of point masses merely had to be replaced by a continuous mass distribution described by a density function $\rho(\boldsymbol{x})$ in a volume V. Conventionally, the entire mass m of the body is attributed to the center of mass. With these conventions, $\boldsymbol{x}(t)$ in Figure 3.3 represents the trajectory of the center of mass, $\boldsymbol{x}_{dm}$, the trajectory of an arbitrary mass element dm of the rigid body.

Coordinate Transformations and Euler Angles. The three-body problem is solved, if the trajectory $\boldsymbol{x}_{dm}(t)$ for each individual mass element of each body is known as a function of time, provided the initial state at an initial epoch t_0 is specified for each mass element.

In principle any coordinate system might be used for this description. An astronomer would prefer to use only the inertial system (and possibly those parallel to it referring to the center of mass of one of the bodies involved), a geodesist would prefer an Earth-fixed system (which inevitably rotates w.r.t. the astronomical systems). The equations of motion are particularly simple in the astronomical system, the description of locations on or within the Earth is particularly simple in the Earth-fixed system. We have to introduce both systems and to establish the transformation between them.

Figure 3.4 illustrates the *geocentric inertial system*, the *Earth-fixed system* and the transformation between the two systems. Admittedly, "geocentric inertial" is not a good designation, because any system attached to a particular point of the Earth shows accelerations w.r.t. the inertial system. By this term we understand a geocentric system which is always parallel to the inertial system. Figure 3.4 shows that the inertial system used here is the ecliptical system referring to a particular epoch (in all applications the realization $J2000.0$ [107] will be used), and that the Earth-fixed system is an equatorial system (in all applications, the ITRF, the International Terrestrial Reference Frame [71], will be used). Figure 3.4 documents that three angles, the so-called Euler angles, are required to perform a coordinate transformation from the geocentric inertial to the Earth-fixed system (and vice versa). Loosely speaking, $\Psi_{♁}$ corresponds to precession (plus nutation) in (ecliptical) longitude, $\varepsilon_{♁}$ to the obliquity of the ecliptic (plus nutation in obliquity), and $\Theta_{♁}$ to sidereal time, if we identify the third axis of the Earth-fixed system

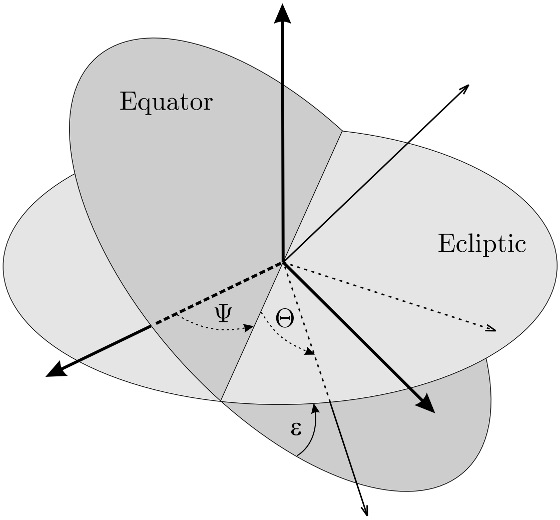


Fig. 3.4. Transformation between the Earth-fixed and the geocentric ecliptical system

with the figure axis of the Earth (pointing approximately to the North pole, to be defined below).

The inertial coordinates $\boldsymbol{r}_{p\mathcal{I}}(t)$ of the geocentric vector $\boldsymbol{r}_p(t)$ of a mass element (or of any other geocentric position vector) may then be computed from the Earth-fixed ones $\boldsymbol{r}_{p\mathcal{F}}(t)$ of the same vector with the following transformation equations:

$$\boldsymbol{r}_{p\mathcal{I}} = \mathbf{R}_3(-\mathit{\Psi}_{\text{♁}})\,\mathbf{R}_1(\varepsilon_{\text{♁}})\,\mathbf{R}_3(-\mathit{\Theta}_{\text{♁}})\,\boldsymbol{r}_{p\mathcal{F}} \overset{\text{def}}{=} \mathbf{T}_{\text{♁}}\,\boldsymbol{r}_{p\mathcal{F}} \ . \tag{3.56}$$

The transformation matrix obviously is a function of all three Euler angles, $\mathbf{T}_{\text{♁}} \overset{\text{def}}{=} \mathbf{T}_{\text{♁}}(\mathit{\Psi}_{\text{♁}}, \varepsilon_{\text{♁}}, \mathit{\Theta}_{\text{♁}})$. Note that it would have been preferable to introduce $\tilde{\varepsilon}_{\text{♁}} \overset{\text{def}}{=} -\varepsilon_{\text{♁}}$ as the inclination of the Earth's equator w.r.t. the ecliptic. In order not to generate confusion, we follow the astronomical conventions in eqn. (3.56) by using the obliquity of the ecliptic w.r.t. the equator (and not the obliquity of the equator w.r.t. the ecliptic) to specify these transformations.

The equation corresponding to a particular mass element of the Moon reads as

$$\boldsymbol{r}_{\wp\mathcal{I}} = \mathbf{R}_3(-\mathit{\Psi}_{\text{☾}})\,\mathbf{R}_1(\varepsilon_{\text{☾}})\,\mathbf{R}_3(-\mathit{\Theta}_{\text{☾}})\,\boldsymbol{r}_{\wp\mathcal{F}} \overset{\text{def}}{=} \mathbf{T}_{\text{☾}}\,\boldsymbol{r}_{\wp\mathcal{F}} \ , \tag{3.57}$$

where the angles $\mathit{\Psi}_{\text{☾}}$, $\varepsilon_{\text{☾}}$, $\mathit{\Theta}_{\text{☾}}$, the transformation matrix $\mathbf{T}_{\text{☾}}$, and the inertial and Moon-fixed coordinates, are defined in an analogous way as the corresponding quantities describing the rotation of the Earth.

Note that eqns. (3.56) and (3.57) are *not* vector equations, but transformation equations for the coordinates of one and the same vector in the two systems.

Euler's Kinematic Equations. In view of the fact that the bodies considered in this section are rigid, the velocity $\dot{\boldsymbol{r}}_{dm}$ of each mass element dm relative to the center of mass of the body is merely due to a rotation about an axis $\boldsymbol{\omega}(t)$ through the center of mass of the body. The velocity of the mass element dm of a body in the geocentric inertial system may therefore be written as

$$\dot{\boldsymbol{r}}_{dm} = \boldsymbol{\omega}(t) \times \boldsymbol{r}_{dm} \ , \tag{3.58}$$

where $\boldsymbol{\omega}$ is the vector of *angular velocity.* Its absolute value $\omega \overset{\text{def}}{=} |\boldsymbol{\omega}|$ is the angular velocity of the body's rotation at time t, the unit vector $\boldsymbol{\omega}/\omega$ is the rotation axis of the body at time t.

Equation (3.58) is a vector equation. It may be evaluated in any coordinate system. In the inertial system the velocities of two mass elements of Earth and Moon may be expressed as follows:

$$\begin{aligned} \dot{\boldsymbol{r}}_{p\mathcal{I}} &= \boldsymbol{\omega}_{\text{♁}\mathcal{I}}(t) \times \boldsymbol{r}_{p\mathcal{I}} \\ \dot{\boldsymbol{r}}_{\wp\mathcal{I}} &= \boldsymbol{\omega}_{\text{☾}\mathcal{I}}(t) \times \boldsymbol{r}_{\wp\mathcal{I}} \ . \end{aligned} \tag{3.59}$$

The velocity of a mass element p of the Earth may, on the other hand, also be computed by taking the time derivative of eqn. (3.56):

$$\begin{aligned}\dot{\boldsymbol{r}}_{p\mathcal{I}} = \Big[\,&\dot{\Psi}_{♁}\,\frac{\partial}{\partial \Psi_{♁}}\,\big\{\mathbf{R}_3(-\Psi_{♁})\big\}\mathbf{R}_1(\varepsilon_{♁})\,\mathbf{R}_3(-\Theta_{♁}) \\ &+\,\dot{\varepsilon}_{♁}\,\mathbf{R}_3(-\Psi_{♁})\,\frac{\partial}{\partial \varepsilon_{♁}}\,\big\{\mathbf{R}_1(\varepsilon_{♁})\big\}\mathbf{R}_3(-\Theta_{♁}) \\ &+\,\dot{\Theta}_{♁}\,\mathbf{R}_3(-\Psi_{♁})\,\mathbf{R}_1(\varepsilon_{♁})\,\frac{\partial}{\partial \Theta_{♁}}\,\big\{\mathbf{R}_3(-\Theta_{♁})\,\big\}\Big]\,\boldsymbol{r}_{p\mathcal{F}}\ .\end{aligned} \tag{3.60}$$

The left-hand sides of eqns. (3.59) (first equation) and (3.60) are identical, which is why the following relationship can be established:

$$\begin{aligned}\boldsymbol{\omega}_{♁\mathcal{I}}(t)\times\boldsymbol{r}_{p\mathcal{I}} = \Big[\,&\dot{\Psi}_{♁}\,\frac{\partial}{\partial \Psi_{♁}}\,\big\{\mathbf{R}_3(-\Psi_{♁})\big\}\mathbf{R}_1(\varepsilon_{♁})\,\mathbf{R}_3(-\Theta_{♁}) \\ &+\,\dot{\varepsilon}_{♁}\,\mathbf{R}_3(-\Psi_{♁})\,\frac{\partial}{\partial \varepsilon_{♁}}\,\big\{\mathbf{R}_1(\varepsilon_{♁})\big\}\mathbf{R}_3(-\Theta_{♁}) \\ &+\,\dot{\Theta}_{♁}\,\mathbf{R}_3(-\Psi_{♁})\,\mathbf{R}_1(\varepsilon_{♁})\,\frac{\partial}{\partial \Theta_{♁}}\,\big\{\mathbf{R}_3(-\Theta_{♁})\,\big\}\Big]\,\boldsymbol{r}_{p\mathcal{F}}\ .\end{aligned} \tag{3.61}$$

Using the coordinate transformation from the inertial to the Earth-fixed system

$$\boldsymbol{r}_{p\mathcal{F}} = \mathbf{R}_3(\Theta_{♁})\,\mathbf{R}_1(-\varepsilon_{♁})\,\mathbf{R}_3(\Psi_{♁})\,\boldsymbol{r}_{p\mathcal{I}} \tag{3.62}$$

to replace $\boldsymbol{r}_{p\mathcal{I}}$ by $\boldsymbol{r}_{p\mathcal{F}}$ on the right-hand side of eqn. (3.61), we obtain the following remarkably simple relation:

$$\begin{aligned}\boldsymbol{\omega}_{♁\mathcal{I}}(t)\times\boldsymbol{r}_{p\mathcal{I}} &= \Big[\dot{\mathbf{T}}_{♁}\,\mathbf{T}_{♁}^{T}\Big]_{\mathcal{I}}\,\boldsymbol{r}_{p\mathcal{F}} \\ &= \left\{\dot{\Psi}_{♁}\begin{pmatrix}0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} + \dot{\varepsilon}_{♁}\begin{pmatrix}0 & 0 & -\sin\Psi_{♁}\\ 0 & 0 & \cos\Psi_{♁}\\ \sin\Psi_{♁} & -\cos\Psi_{♁} & 0\end{pmatrix}\right. \\ &\left.+\,\dot{\Theta}_{♁}\begin{pmatrix}0 & -\cos\varepsilon_{♁} & \sin\varepsilon_{♁}\cos\Psi_{♁}\\ \cos\varepsilon_{♁} & 0 & \sin\varepsilon_{♁}\sin\Psi_{♁}\\ -\sin\varepsilon_{♁}\cos\Psi_{♁} & -\sin\varepsilon_{♁}\sin\Psi_{♁} & 0\end{pmatrix}\right\}\boldsymbol{r}_{p\mathcal{I}}\ ,\end{aligned} \tag{3.63}$$

where $[\ldots]_{\mathcal{I}}$ indicates that the expression refers to the inertial system.

In view of the fact that the left-hand side of equation (3.63) may be written in the following matrix form

$$\boldsymbol{\omega}_{♁\mathcal{I}}(t)\times\boldsymbol{r}_{p\mathcal{I}} = \Big[\dot{\mathbf{T}}_{♁}\mathbf{T}_{♁}^{T}\Big]_{\mathcal{I}}\,\boldsymbol{r}_{p\mathcal{I}} = \begin{pmatrix}0 & -\omega_{♁\mathcal{I}_3} & +\omega_{♁\mathcal{I}_2}\\ +\omega_{♁\mathcal{I}_3} & 0 & -\omega_{♁\mathcal{I}_1}\\ -\omega_{♁\mathcal{I}_2} & +\omega_{♁\mathcal{I}_1} & 0\end{pmatrix}\boldsymbol{r}_{p\mathcal{I}}\ , \tag{3.64}$$

we obtain the equation relating the components of the angular velocity vector in the inertial system to the first derivatives of the Euler angles:

$$\boldsymbol{\omega}_{♁\mathcal{I}}(t) = \begin{pmatrix} 0 & -\cos\varPsi_{♁} & -\sin\varepsilon_{♁}\sin\varPsi_{♁} \\ 0 & -\sin\varPsi_{♁} & +\sin\varepsilon_{♁}\cos\varPsi_{♁} \\ 1 & 0 & \cos\varepsilon_{♁} \end{pmatrix} \begin{pmatrix} \dot{\varPsi}_{♁} \\ \dot{\varepsilon}_{♁} \\ \dot{\varTheta}_{♁} \end{pmatrix} . \tag{3.65}$$

The inverse transformation is given by

$$\begin{pmatrix} \dot{\varPsi}_{♁} \\ \dot{\varepsilon}_{♁} \\ \dot{\varTheta}_{♁} \end{pmatrix} = \begin{pmatrix} \sin\varPsi_{♁}\cot\varepsilon_{♁} & -\cos\varPsi_{♁}\cot\varepsilon_{♁} & 1 \\ -\cos\varPsi_{♁} & -\sin\varPsi_{♁} & 0 \\ -\sin\varPsi_{♁}\csc\varepsilon_{♁} & +\cos\varPsi_{♁}\csc\varepsilon_{♁} & 0 \end{pmatrix} \boldsymbol{\omega}_{♁\mathcal{I}}(t) \overset{\text{def}}{=} \mathbf{X}_{♁\mathcal{I}}\,\boldsymbol{\omega}_{♁\mathcal{I}}(t) \; . \tag{3.66}$$

Equations (3.59) and (3.60) were used to establish the relationship between the Cartesian coordinates of the angular velocity vector of Earth rotation in the inertial coordinate system and the first time derivatives of the Euler angles, $\dot{\varPsi}_{♁}$, $\dot{\varepsilon}_{♁}$, and $\dot{\varTheta}_{♁}$. The same equations may be used to establish the relationship between the Cartesian coordinates of the same vector in the Earth-fixed system and the first derivatives of the Euler angles. The calculations are straight forward and lead to the following result:

$$\boldsymbol{\omega}_{♁\mathcal{F}}(t) = \begin{pmatrix} -\sin\varepsilon_{♁}\sin\varTheta_{♁} & -\cos\varTheta_{♁} & 0 \\ -\sin\varepsilon_{♁}\cos\varTheta_{♁} & +\sin\varTheta_{♁} & 0 \\ \cos\varepsilon_{♁} & 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{\varPsi}_{♁} \\ \dot{\varepsilon}_{♁} \\ \dot{\varTheta}_{♁} \end{pmatrix} . \tag{3.67}$$

The inverse transformation is easily derived from the above equation:

$$\begin{pmatrix} \dot{\varPsi}_{♁} \\ \dot{\varepsilon}_{♁} \\ \dot{\varTheta}_{♁} \end{pmatrix} = \begin{pmatrix} -\sin\varTheta_{♁}\csc\varepsilon_{♁} & -\cos\varTheta_{♁}\csc\varepsilon_{♁} & 0 \\ -\cos\varTheta_{♁} & +\sin\varTheta_{♁} & 0 \\ \sin\varTheta_{♁}\cot\varepsilon_{♁} & +\cos\varTheta_{♁}\cot\varepsilon_{♁} & 1 \end{pmatrix} \boldsymbol{\omega}_{♁\mathcal{F}}(t) \overset{\text{def}}{=} \mathbf{X}_{♁\mathcal{F}}\,\boldsymbol{\omega}_{♁\mathcal{F}}(t) \; . \tag{3.68}$$

Equations (3.65, 3.66) and (3.67, 3.68) also are referred to as *Euler's kinematic equations* of Earth rotation. We derived them in a purely algebraic way. It is also possible to give a geometrical derivation by projecting the angular velocities $\dot{\varPsi}_{♁}$, $\dot{\varepsilon}_{♁}$, and $\dot{\varTheta}_{♁}$ on the resp. coordinate axes.

For later use we note the result, which was established as a side issue, so to speak *en passant* in this section: It is obviously possible to express the matrix $\dot{\mathbf{T}}_{♁}\mathbf{T}^T_{♁}$ either through the components of the angular velocity vector in the inertial system, eqn. (3.64), or to the Earth-fixed system (analogous to eqn. (3.64) related to the Earth-fixed system):

$$\begin{aligned} \left[\dot{\mathbf{T}}_{♁}\mathbf{T}^T_{♁}\right]_{\mathcal{I}} &= \begin{pmatrix} 0 & -\omega_{♁\mathcal{I}_3} & +\omega_{♁\mathcal{I}_2} \\ +\omega_{♁\mathcal{I}_3} & 0 & -\omega_{♁\mathcal{I}_1} \\ -\omega_{♁\mathcal{I}_2} & +\omega_{♁\mathcal{I}_1} & 0 \end{pmatrix} , \\ \left[\dot{\mathbf{T}}_{♁}\mathbf{T}^T_{♁}\right]_{\mathcal{F}} &= \begin{pmatrix} 0 & -\omega_{♁\mathcal{F}_3} & +\omega_{♁\mathcal{F}_2} \\ +\omega_{♁\mathcal{F}_3} & 0 & -\omega_{♁\mathcal{F}_1} \\ -\omega_{♁\mathcal{F}_2} & +\omega_{♁\mathcal{F}_1} & 0 \end{pmatrix} . \end{aligned} \tag{3.69}$$

Similar relationships may be derived for the Euler angles for the rotation of the Moon. They follow from eqns. (3.65, 3.67) by replacing index ♁ by ☾.

Angular Momentum Vector and Inertia Tensor. The angular momentum vector of a rigid body w.r.t. a center of mass coordinate system is defined as

$$\boldsymbol{h} = \int\limits_V \rho(\boldsymbol{r}_{dm})\ \boldsymbol{r}_{dm} \times \dot{\boldsymbol{r}}_{dm}\ dV = \int\limits_V \rho(\boldsymbol{r}_{dm})\ \boldsymbol{r}_{dm} \times \left(\boldsymbol{\omega} \times \boldsymbol{r}_{dm}\right)\ dV\ . \qquad (3.70)$$

Using the well-known relation

$$\boldsymbol{r}_{dm} \times (\boldsymbol{\omega} \times \boldsymbol{r}_{dm}) = r_{dm}^2\ \boldsymbol{\omega} - (\boldsymbol{\omega}\ \boldsymbol{r}_{dm})\ \boldsymbol{r}_{dm} \qquad (3.71)$$

we obtain the following equation for the angular momentum vector of the rigid body:

$$\boldsymbol{h} = \int\limits_V \rho(\boldsymbol{r}_{dm})\ \left\{ r_{dm}^2\ \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \boldsymbol{r}_{dm})\ \boldsymbol{r}_{dm} \right\}\ dV\ . \qquad (3.72)$$

The expression shows that the Cartesian components of vector $\boldsymbol{h}$ are linear functions of the components of the angular velocity vector $\boldsymbol{\omega}$. It may therefore be written in the following elegant way:

$$\boldsymbol{h} = \left\{ \int\limits_V \rho(\boldsymbol{r}_{dm})\ \left[r_{dm}^2\ \mathbf{E} - \boldsymbol{r}_{dm} \otimes \boldsymbol{r}_{dm} \right]\ dV \right\} \boldsymbol{\omega} \overset{\text{def}}{=} \mathbf{I}\ \boldsymbol{\omega}\ , \qquad (3.73)$$

where $\boldsymbol{r}_{dm} \otimes \boldsymbol{r}_{dm}$ is the outer product (or tensor product) of the vector (or tensor of rank 1) $\boldsymbol{r}_{dm}$ with itself, $\mathbf{E}$ is the unit tensor of rank two, $\mathbf{I}$ is the *inertia tensor*, a tensor of rank two, as well. The inertia tensor $\mathbf{I}$ of a body is thus defined by:

$$\mathbf{I} = \int\limits_V \rho(\boldsymbol{r}_{dm})\ \left[r_{dm}^2\ \mathbf{E} - \boldsymbol{r}_{dm} \otimes \boldsymbol{r}_{dm} \right]\ dV\ . \qquad (3.74)$$

Transformation Law for the Inertia Tensor, Principal Moments of Inertia. The inertia tensor may be expressed in any coordinate system. The tensor notation then becomes an ordinary matrix notation. Let us compute the inertia tensor of the Earth in the inertial system as an example:

$$\mathbf{I}_{♁\mathcal{I}} = \int\limits_{V_♁} \rho(\boldsymbol{r}_{p\mathcal{I}})\ \left[r_p^2\ \mathbf{E} - \boldsymbol{r}_{p\mathcal{I}} \otimes \boldsymbol{r}_{p\mathcal{I}} \right]\ dV_♁\ . \qquad (3.75)$$

As the Earth performs rather complicated rotations in the inertial system, the inertia tensor, expressed in the inertial system, is a quite complicated function of time, as well. Wherever possible, one should therefore refer the inertia tensor to an Earth-fixed system:

$$\mathbf{I}_{\text{♁}\mathcal{F}} = \int\limits_{V_{\text{♁}}} \rho(\boldsymbol{r}_{p\mathcal{F}}) \left[r^2_{p\mathcal{F}}\ \mathbf{E} - \boldsymbol{r}_{p\mathcal{F}} \otimes \boldsymbol{r}_{p\mathcal{F}} \right] dV_{\text{♁}}\ . \tag{3.76}$$

In view of the transformation equations (3.56) one easily establishes the transformation law for the inertia tensor in the inertial and Earth-fixed coordinate system:

$$\begin{aligned} \mathbf{I}_{\text{♁}\mathcal{I}} &= \mathbf{T}_{\text{♁}}\, \mathbf{I}_{\text{♁}\mathcal{F}}\, \mathbf{T}^T_{\text{♁}} \\ &= \mathbf{R}_3(-\Psi_{\text{♁}})\, \mathbf{R}_1(\varepsilon_{\text{♁}})\, \mathbf{R}_3(-\Theta_{\text{♁}})\, \mathbf{I}_{\text{♁}\mathcal{F}}\, \mathbf{R}_3(\Theta_{\text{♁}})\, \mathbf{R}_1(-\varepsilon_{\text{♁}})\, \mathbf{R}_3(\Psi_{\text{♁}})\ . \end{aligned} \tag{3.77}$$

By definition, $\mathbf{I}_{\text{♁}\mathcal{F}}$ is constant in time for a rigid Earth. In general, the matrix will be fully populated, however.

So far, we have not put any requirements on the Earth-fixed system other than letting its origin coincide with the center of mass. We are now in a position to introduce a particularly well-suited Earth-fixed system, namely the one with respect to which the inertia tensor becomes diagonal. Is it always possible to find such a system for any given density distribution? The answer is given by the fact that it is always possible to perform a main-axis transformation (a rotation in our case) $\tilde{\boldsymbol{r}}_{\text{♁}\mathcal{F}} = \tilde{\mathbf{A}}\, \boldsymbol{r}_{\text{♁}\mathcal{F}}$ such that $\tilde{\mathbf{I}}_{\text{♁}\mathcal{F}} = \tilde{\mathbf{A}}\, \mathbf{I}_{\text{♁}\mathcal{F}}\, \tilde{\mathbf{A}}^T$ becomes diagonal. The technique consists of finding the *eigenvalues* and then the *eigenvectors* of the matrix $\mathbf{I}_{\text{♁}\mathcal{F}}$. We may thus assume that the coordinate axes of the Earth-fixed system coincide with the *principal axes of inertia* and that the inertia tensor in this system may be written as

$$\mathbf{I}_{\text{♁}\mathcal{F}} \overset{\text{def}}{=} \begin{pmatrix} I_{\text{♁}1} & 0 & \\ 0 & I_{\text{♁}2} & 0 \\ 0 & 0 & I_{\text{♁}3} \end{pmatrix} = \begin{pmatrix} A_{\text{♁}} & 0 & 0 \\ 0 & B_{\text{♁}} & 0 \\ 0 & 0 & C_{\text{♁}} \end{pmatrix} \approx \begin{pmatrix} A_{\text{♁}} & 0 & 0 \\ 0 & A_{\text{♁}} & 0 \\ 0 & 0 & C_{\text{♁}} \end{pmatrix}\ , \tag{3.78}$$

where the quantities $I_{\text{♁}1} = A_{\text{♁}}\ ,\ I_{\text{♁}2} = B_{\text{♁}}\ ,\ I_{\text{♁}3} = C_{\text{♁}}\ ,\ A_{\text{♁}} < B_{\text{♁}} < C_{\text{♁}}$ are called the *principal moments of inertia.*

If the inertia tensor of a celestial body becomes diagonal (as given by eqns. (3.78) for the Earth), we call the underlying coordinate system the *coordinate system of the PAI (Principal Axes of Inertia)* or simply *PAI-system.*

Their numerical values are listed in Table II- 2.1. In the case of the Earth, rotational symmetry $B \approx A$ is an excellent approximation (as indicated in the above equation). The principal axis corresponding to the maximum moment of inertia, also called the *figure axis* of the planet, approximately points to the North pole. Similar equations result for the inertia tensor of the Moon, except for the approximation in equation (3.78) which is not justified in the case of the Moon.

In the PAI-system the elements of the inertia tensor are defined as follows (see eqn. (3.76) and take into account eqn. (3.78)):

$$\begin{aligned} \mathbf{I}_{\mathcal{F}_{11}} &= A = \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\,(r^2_{\mathcal{F}_2} + r^2_{\mathcal{F}_3})\,dV \\ \mathbf{I}_{\mathcal{F}_{22}} &= B = \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\,(r^2_{\mathcal{F}_1} + r^2_{\mathcal{F}_3})\,dV \\ \mathbf{I}_{\mathcal{F}_{33}} &= C = \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\,(r^2_{\mathcal{F}_1} + r^2_{\mathcal{F}_2})\,dV \\ \mathbf{I}_{\mathcal{F}_{ik}} &\quad\;\, = 0 \quad \text{for} \;\; i \neq k \; . \end{aligned} \tag{3.79}$$

These equations allow the establishment of the following useful relations:

$$\begin{aligned} \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\; r^2\; dV &= \tfrac{1}{2}\;(A+B+C) \\ \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\; r_1^2\; dV &= \tfrac{1}{2}\;(A+B+C) - A \\ \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\; r_2^2\; dV &= \tfrac{1}{2}\;(A+B+C) - B \\ \int\limits_V \rho(\boldsymbol{r}_{\wp\mathcal{F}})\; r_3^2\; dV &= \tfrac{1}{2}\;(A+B+C) - C\; . \end{aligned} \tag{3.80}$$

We have established the transformation equations and the PAI-system for the Earth. Equations of the same type may be developed for the Moon. Apart from replacing the index "♁" by "☾" they are identical with the corresponding relations for the Earth. It is interesting to note (see Table II- 2.1) that (as opposed to the Earth) rotational symmetry is not an appropriate approximation in the case of the Moon's principal moments of inertia. Note that (as in the case of the Earth) the axis of maximum moment of inertia closely coincides with the rotation axis of the Moon, and that the axis of minimum moment of inertia approximately points toward the Earth.

3.3.3 The Equations of Motion in the Inertial System

As in the case of the N-body problem the equations of motion first are established in the inertial system. We set up one equation of motion for each mass element p of the Earth and one for each mass element $\wp$ of the Moon. The Sun is treated as one point mass. Afterwards, the equations of motion are derived for the center of mass of Earth and Moon and for the angular momenta associated with Earth and Moon. Gravitational forces between mass elements of different bodies of course have to be taken into account. In addition, other forces between the mass elements of one and the same body may be considered, as long as they act along the line between the two mass elements, and as long as the third Newtonian axiom (see Table 2.2) holds:

$$\begin{aligned}
p\,\ddot{\boldsymbol{x}}_p &= -\,G\,m_\odot\,p\,\frac{\boldsymbol{x}_p-\boldsymbol{x}_\odot}{|\boldsymbol{x}_p-\boldsymbol{x}_\odot|^3} \;-\; G\,p\int\limits_{V_{\text{☾}}}\rho(\boldsymbol{x}_\wp)\,\frac{\boldsymbol{x}_p-\boldsymbol{x}_\wp}{|\boldsymbol{x}_p-\boldsymbol{x}_\wp|^3}\,dV_{\text{☾}} \;+\int\limits_{V_{\text{♁}}}\boldsymbol{f}_{p,p'}\,dV_{\text{♁}}\\
\wp\,\ddot{\boldsymbol{x}}_\wp &= -\,G\,m_\odot\,\wp\,\frac{\boldsymbol{x}_\wp-\boldsymbol{x}_\odot}{|\boldsymbol{x}_\wp-\boldsymbol{x}_\odot|^3} \;-\; G\,\wp\int\limits_{V_{\text{♁}}}\rho(\boldsymbol{x}_p)\,\frac{\boldsymbol{x}_\wp-\boldsymbol{x}_p}{|\boldsymbol{x}_\wp-\boldsymbol{x}_p|^3}\,dV_{\text{♁}} \;+\int\limits_{V_{\text{☾}}}\boldsymbol{f}_{\wp,\wp'}\,dV_{\text{☾}}\\
\ddot{\boldsymbol{x}}_\odot &= -\,G\int\limits_{V_{\text{♁}}}\rho(\boldsymbol{x}_p)\,\frac{\boldsymbol{x}_\odot-\boldsymbol{x}_p}{|\boldsymbol{x}_\odot-\boldsymbol{x}_p|^3}\,dV_{\text{♁}} \;-\; G\int\limits_{V_{\text{☾}}}\rho(\boldsymbol{x}_\wp)\,\frac{\boldsymbol{x}_\odot-\boldsymbol{x}_\wp}{|\boldsymbol{x}_\odot-\boldsymbol{x}_\wp|^3}\,dV_{\text{☾}}\ ,
\end{aligned} \tag{3.81}$$

where $\boldsymbol{f}_{p,p'}$ and $\boldsymbol{f}_{\wp,\wp'}$ are internal forces acting from one mass element of a body to another mass element of the same body.

The equations of motion are vector equations. They may, however, also be interpreted as coordinate equations referring to the inertial system. In order to reduce the formalism, the index specifying the coordinate system will be omitted, wherever feasible.

The integrals over the gravitational and non-gravitational forces within the same body have to be performed cautiously: The volume element occupied by the mass elements p and $\wp$ on the left-hand side of the above equations must be left out in the integration over the volumes of bodies on the right-hand sides.

The equation of motion for the Sun is already in a useable form. Using the decomposition $\boldsymbol{x}_p = \boldsymbol{x}_{\text{♁}} + \boldsymbol{r}_p$, where $\boldsymbol{x}_{\text{♁}}$ is the position vector of the Earth's center of mass and the analogous decomposition for the position vector of the Moon's mass elements, one obtains by integrating over the equations of motion of all mass elements of Earth and Moon, respectively, the equations of motion for the centers of mass of the two bodies:

$$\begin{aligned}
\ddot{\boldsymbol{x}}_{♁} &= -\,G\,m_{\odot} \int\limits_{V_{♁}} \frac{\rho(\boldsymbol{x}_p)}{M}\,\frac{\boldsymbol{x}_p - \boldsymbol{x}_{\odot}}{|\boldsymbol{x}_p - \boldsymbol{x}_{\odot}|^3}\;dV_{♁} \\
&\quad -\,G\,m_{\leftmoon} \int\limits_{V_{♁}} \frac{\rho(\boldsymbol{x}_p)}{M} \int\limits_{V_{\leftmoon}} \frac{\rho(\boldsymbol{x}_{\wp})}{m_{\leftmoon}}\,\frac{\boldsymbol{x}_p - \boldsymbol{x}_{\wp}}{|\boldsymbol{x}_p - \boldsymbol{x}_{\wp}|^3}\;dV_{\leftmoon}\;dV_{♁} \\
\ddot{\boldsymbol{x}}_{\leftmoon} &= -\,G\,m_{\odot} \int\limits_{V_{\leftmoon}} \frac{\rho(\boldsymbol{x}_{\wp})}{m_{\leftmoon}}\,\frac{\boldsymbol{x}_{\wp} - \boldsymbol{x}_{\odot}}{|\boldsymbol{x}_{\wp} - \boldsymbol{x}_{\odot}|^3}\;dV_{\leftmoon} \\
&\quad -\,GM \int\limits_{V_{\leftmoon}} \frac{\rho(\boldsymbol{x}_{\wp})}{m_{\leftmoon}} \int\limits_{V_{♁}} \frac{\rho(\boldsymbol{x}_p)}{M}\,\frac{\boldsymbol{x}_{\wp} - \boldsymbol{x}_p}{|\boldsymbol{x}_{\wp} - \boldsymbol{x}_p|^3}\;dV_{♁}\;dV_{\leftmoon} \\
\ddot{\boldsymbol{x}}_{\odot} &= -\,GM \int\limits_{V_{♁}} \frac{\rho(\boldsymbol{x}_p)}{M}\,\frac{\boldsymbol{x}_{\odot} - \boldsymbol{x}_p}{|\boldsymbol{x}_{\odot} - \boldsymbol{x}_p|^3}\;dV_{♁} \;-\; G\,m_{\leftmoon} \int\limits_{V_{\leftmoon}} \frac{\rho(\boldsymbol{x}_{\wp})}{m_{\leftmoon}}\,\frac{\boldsymbol{x}_{\odot} - \boldsymbol{x}_{\wp}}{|\boldsymbol{x}_{\odot} - \boldsymbol{x}_{\wp}|^3}\;dV_{\leftmoon}\;.
\end{aligned} \tag{3.82}$$

The (double) integral $\int\limits_{V_{♁}}\int\limits_{V'_{♁}} \boldsymbol{f}_{p,p'}\,dV_{♁}\,dV'_{♁}$ (and the corresponding integral for the Moon) are zero by virtue of Newton's third axiom. The above result was obtained after division by the total mass of the bodies (observe the definition (3.55) for a body's center of mass).

In analogy to the derivation of the heliocentric equations of motion of the planetary system, the equations of motion for the centers of mass of Sun and Moon will be referred to the center of mass of the Earth as central body. Let us recall here the notations shown in Figure 3.2 and introduce shorter notations for the densities of Earth and Moon:

$$\begin{aligned}
\boldsymbol{r}_{\odot} &= \boldsymbol{x}_{\odot} - \boldsymbol{x}_{♁} \\
\boldsymbol{r}_{\leftmoon} &= \boldsymbol{x}_{\leftmoon} - \boldsymbol{x}_{♁} \\
\boldsymbol{x}_{\wp} - \boldsymbol{x}_p &= \boldsymbol{r}_{\wp} - \boldsymbol{r}_p \\
&= \boldsymbol{r}_{\leftmoon} + \boldsymbol{r}_{\leftmoon\wp} - \boldsymbol{r}_p \\
\ldots &= \ldots \\
\\
\rho_p &\stackrel{\text{def}}{=} \rho(\boldsymbol{x}_p) \\
\rho_{\wp} &\stackrel{\text{def}}{=} \rho(\boldsymbol{x}_{\wp}) \\
\rho_{p_r} &= \frac{\rho(\boldsymbol{x}_p)}{M} \\
\rho_{\wp_r} &= \frac{\rho(\boldsymbol{x}_{\leftmoon\wp})}{m_{\leftmoon}}\;.
\end{aligned} \tag{3.83}$$

The relative densities ρ_{p_r} and $\rho_{\wp_r}$ express the density as mass per volume, where the total mass of the body serves as unit of mass.

Let us refer the equations for the geocentric motion to the geocentric inertial system introduced previously. The equations for the geocentric motion of Moon and Sun are obtained by subtracting the equation for the motion of the Earth's center of mass from the equations for the centers of mass of the other two bodies in eqn. (3.82):

$$\begin{aligned}
\ddot{\boldsymbol{r}}_{\leftmoon} &= -\,G\,(M+m_{\leftmoon}) \int\limits_{V_{\leftmoon}} \int\limits_{V_{\earth}} \rho_{p_r}\rho_{\wp_r}\, \frac{\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p|^3}\; dV_{\earth}\; dV_{\leftmoon} \\
&\qquad -\,G\,m_{\odot} \left\{ \int\limits_{V_{\leftmoon}} \rho_{\wp_r}\, \frac{\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_{\odot}|^3}\; dV_{\leftmoon} \;+ \int\limits_{V_{\earth}} \rho_{p_r}\, \frac{\boldsymbol{r}_{\odot}-\boldsymbol{r}_p}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_p|^3}\; dV_{\earth} \right\} \\
\ddot{\boldsymbol{r}}_{\odot} &= -\,G\,(m_{\odot}+M) \int\limits_{V_{\earth}} \rho_{p_r}\, \frac{\boldsymbol{r}_{\odot}-\boldsymbol{r}_p}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_p|^3}\; dV_{\earth} \\
&\qquad -\,G\,m_{\leftmoon} \int\limits_{V_{\leftmoon}} \rho_{\wp_r} \left\{ \frac{\boldsymbol{r}_{\odot}-\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\leftmoon\wp}}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\leftmoon\wp}|^3} + \int\limits_{V_{\earth}} \rho_{p_r}\, \frac{\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p|^3}\; dV_{\earth} \right\} dV_{\leftmoon}\ .
\end{aligned} \tag{3.84}$$

The right-hand sides of these equations may be written as gradients of scalar functions. In the equation for the Moon the gradients refer to the geocentric coordinates of the Moon, in the equation for the Sun to those of the Sun. This is indicated by the indices "$\odot$" and "$\leftmoon$".

$$\begin{aligned}
\ddot{\boldsymbol{r}}_{\leftmoon} &= G\,(M+m_{\leftmoon})\,\nabla_{\leftmoon} \left\{ \int\limits_{V_{\leftmoon}} \int\limits_{V_{\earth}} \frac{\rho_{p_r}\,\rho_{\wp_r}}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p|}\; dV_{\earth}\; dV_{\leftmoon} \right\} \\
&\qquad +\,G\,m_{\odot}\,\nabla_{\leftmoon} \left\{ \int\limits_{V_{\leftmoon}} \frac{\rho_{\wp_r}}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_{\odot}|}\; dV_{\leftmoon} \;- \int\limits_{V_{\earth}} \rho_{p_r}\, \frac{(\boldsymbol{r}_{\odot}-\boldsymbol{r}_p)\cdot\boldsymbol{r}_{\leftmoon}}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_p|^3}\; dV_{\earth} \right\} \\
\ddot{\boldsymbol{r}}_{\odot} &= G\,(m_{\odot}+M)\,\nabla_{\odot} \int\limits_{V_{\earth}} \frac{\rho_{p_r}}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_p|}\; dV_{\earth} \;+\; G\,m_{\leftmoon}\,\nabla_{\odot} \left\{ \int\limits_{V_{\leftmoon}} \frac{\rho_{\wp_r}}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\leftmoon\wp}|}\, dV_{\leftmoon} \right. \\
&\qquad \left. -\int\limits_{V_{\leftmoon}} \rho_{\wp_r} \int\limits_{V_{\earth}} \rho_{p_r}\, \frac{(\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p)\cdot\boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p|^3}\; dV_{\earth}\; dV_{\leftmoon} \right\}\ .
\end{aligned} \tag{3.85}$$

Equations (3.85) are the equations of orbital motion for a three-body problem with two bodies of finite size and one point mass. They reduce to the corresponding equations (3.22) for point masses, if the mass distribution is spherically symmetric. The generalized two-body motion is described by the first terms on the right-hand sides of eqns. (3.85), the generalized perturbation terms are the second terms on the right-hand sides. Obviously the above equations are not yet convenient for the solution of the problem. Integrations over the entire volume of the body are required at each integration step. Approximate and easy to use expressions for the integrals will be developed in section 3.3.5.

The equations for the rotation of Earth and Moon are obtained by multiplying the equations for the mass elements for Earth and Moon in eqns. (3.81) vectorially (from the left) with the geocentric and selenocentric position vectors $\boldsymbol{r}_p$ and $\boldsymbol{r}_{\text{☾}\wp}$ and by integrating over the entire volume of the bodies. Representing the vectors $\boldsymbol{x}_p$ and $\boldsymbol{x}_{\text{☾}\wp}$ on the left-hand side as the sums $\boldsymbol{x}_p = \boldsymbol{x}_{\text{♁}} + \boldsymbol{r}_p$ and $\boldsymbol{x}_\wp = \boldsymbol{x}_{\text{☾}} + \boldsymbol{r}_{\text{☾}\wp}$ of the centers's of mass vectors and the geo- and selenocentric position vectors of the mass elements, and taking into account the center of mass definition (3.55) for rigid bodies, the following equations for the angular momenta of the two bodies result:

$$
\begin{aligned}
\int\limits_{V_{\text{♁}}} \rho_p \, \boldsymbol{r}_p \times \ddot{\boldsymbol{r}}_p \, dV_{\text{♁}} &= G\, m_\odot \int\limits_{V_{\text{♁}}} \rho_p \, \frac{\boldsymbol{r}_p \times \boldsymbol{r}_\odot}{|\boldsymbol{r}_\odot - \boldsymbol{r}_p|^3} \, dV_{\text{♁}} \\
&\quad + G \int\limits_{V_{\text{♁}}} \int\limits_{V_{\text{☾}}} \rho_p \, \rho_\wp \, \frac{\boldsymbol{r}_p \times (\boldsymbol{r}_{\text{☾}} + \boldsymbol{r}_{\text{☾}\wp})}{|\boldsymbol{r}_p - \boldsymbol{r}_{\text{☾}} - \boldsymbol{r}_{\text{☾}\wp}|^3} \, dV_{\text{☾}} \, dV_{\text{♁}} \\
\int\limits_{V_{\text{☾}}} \rho_\wp \, \boldsymbol{r}_{\text{☾}\wp} \times \ddot{\boldsymbol{r}}_{\text{☾}\wp} \, dV_{\text{☾}} &= G\, m_\odot \int\limits_{V_{\text{☾}}} \rho_\wp \, \frac{\boldsymbol{r}_{\text{☾}\wp} \times (\boldsymbol{r}_\odot - \boldsymbol{r}_{\text{☾}})}{|\boldsymbol{r}_\odot - \boldsymbol{r}_{\text{☾}} - \boldsymbol{r}_{\text{☾}\wp}|^3} \, dV_{\text{☾}} \\
&\quad + G \int\limits_{V_{\text{♁}}} \int\limits_{V_{\text{☾}}} \rho_p \, \rho_\wp \, \frac{\boldsymbol{r}_{\text{☾}\wp} \times (\boldsymbol{r}_p - \boldsymbol{r}_{\text{☾}})}{|\boldsymbol{r}_{\text{☾}} + \boldsymbol{r}_{\text{☾}\wp} - \boldsymbol{r}_p|^3} \, dV_{\text{☾}} \, dV_{\text{♁}} \ .
\end{aligned}
\tag{3.86}
$$

A comparison of the left-hand side of the above equations with the definition (3.70) for the angular momentum of a celestial body reveals, that these may be written as the time derivative of the bodies' angular momenta:

$$
\begin{aligned}
\dot{\boldsymbol{h}}_{\text{♁}} &= \int\limits_{V_{\text{♁}}} \rho_p \, \boldsymbol{r}_p \times \ddot{\boldsymbol{r}}_p \, dV_{\text{♁}} \\
\dot{\boldsymbol{h}}_{\text{☾}} &= \int\limits_{V_{\text{☾}}} \rho_\wp \, \boldsymbol{r}_{\text{☾}\wp} \times \ddot{\boldsymbol{r}}_{\text{☾}\wp} \, dV_{\text{☾}} \ .
\end{aligned}
\tag{3.87}
$$

Taking into account that the gravitational attractions on the right-hand side may be written as gradients of scalar functions, we obtain the equations for the rotation of the Earth and the Moon:

$$\begin{aligned}
\dot{\boldsymbol{h}}_{\earth} &= + G\, m_{\odot} \int\limits_{V_{\earth}} \rho_p \, \nabla_{\odot} \left\{ \frac{1}{|\boldsymbol{r}_{\odot} - \boldsymbol{r}_p|} \right\} \times \boldsymbol{r}_{\odot} \; dV_{\earth} \\
&\quad + G \int\limits_{V_{\earth}} \int\limits_{V_{\leftmoon}} \rho_p \, \rho_{\wp} \nabla_{\leftmoon} \left\{ \frac{1}{|\boldsymbol{r}_p - \boldsymbol{r}_{\leftmoon} - \boldsymbol{r}_{\leftmoon\wp}|} \right\} \times (\boldsymbol{r}_{\leftmoon\wp} + \boldsymbol{r}_{\leftmoon}) \;\; dV_{\leftmoon} \;\; dV_{\earth} \\
\dot{\boldsymbol{h}}_{\leftmoon} &= + G\, m_{\odot} \int\limits_{V_{\leftmoon}} \rho_{\wp} \; \nabla_{\odot\leftmoon} \left\{ \frac{1}{|\boldsymbol{r}_{\odot} - \boldsymbol{r}_{\leftmoon} - \boldsymbol{r}_{\leftmoon\wp}|} \right\} \times (\boldsymbol{r}_{\odot} - \boldsymbol{r}_{\leftmoon}) \;\; dV_{\leftmoon} \\
&\quad - G \int\limits_{V_{\earth}} \int\limits_{V_{\leftmoon}} \rho_p \, \rho_{\wp} \nabla_{\leftmoon} \left\{ \frac{1}{|\boldsymbol{r}_p - \boldsymbol{r}_{\leftmoon} - \boldsymbol{r}_{\leftmoon\wp}|} \right\} \times (\boldsymbol{r}_p - \boldsymbol{r}_{\leftmoon}) \;\; dV_{\leftmoon} \;\; dV_{\earth} \; .
\end{aligned} \tag{3.88}$$

Note that the index of the gradient symbol indicates with respect to which coordinates the gradient has to be taken.

As the *torque* acting on a body through a force $\boldsymbol{f}$ is defined as

$$\boldsymbol{\ell} \overset{\text{def}}{=} \int\limits_{V} \rho \; \boldsymbol{r}_{dm} \times \boldsymbol{f} \; dV \;\; , \tag{3.89}$$

eqns. (3.88) express the physical law that the change of angular momentum of a body is due to (and equal to) the sum of external torques acting on the body:

$$\begin{aligned}
\dot{\boldsymbol{h}}_{\earth} &= \boldsymbol{\ell}_{\odot\earth} + \boldsymbol{\ell}_{\leftmoon\earth} \\
\dot{\boldsymbol{h}}_{\leftmoon} &= \boldsymbol{\ell}_{\odot\leftmoon} + \boldsymbol{\ell}_{\earth\leftmoon} \;\; ,
\end{aligned} \tag{3.90}$$

where $\boldsymbol{\ell}_{\odot\earth}$, $\boldsymbol{\ell}_{\leftmoon\earth}$ are the torques exerted by Sun and Moon on the Earth, $\boldsymbol{\ell}_{\odot\leftmoon}$, $\boldsymbol{\ell}_{\earth\leftmoon}$ those exerted by Sun and Earth on the Moon.

In the previous section the relations between the angular momentum $\boldsymbol{h}_{\dots}$ and the angular velocity vector $\boldsymbol{\omega}_{\dots}$, and between the components of the angular velocity vector and the first derivatives of the Euler angles were established (see eqns. (3.73) and (3.65)). From these equations we may directly establish the relation between the angular momentum vector and the Euler angles:

$$\begin{aligned} \begin{pmatrix} \dot{\Psi}_{♁} \\ \dot{\varepsilon}_{♁} \\ \dot{\Theta}_{♁} \end{pmatrix} &= \mathbf{X}_{♁\mathcal{I}}\, \mathbf{I}^{-1}_{♁\mathcal{I}}\, \boldsymbol{h}_{♁\mathcal{I}} = \mathbf{X}_{♁\mathcal{I}}\, \mathbf{T}_{♁}\, \mathbf{I}^{-1}_{♁\mathcal{F}}\, \mathbf{T}^{T}_{♁}\, \boldsymbol{h}_{♁\mathcal{I}} \overset{\text{def}}{=} \mathbf{Y}_{♁\mathcal{I}}\, \boldsymbol{h}_{♁\mathcal{I}} \\ \begin{pmatrix} \dot{\Psi}_{☾} \\ \dot{\varepsilon}_{☾} \\ \dot{\Theta}_{☾} \end{pmatrix} &= \mathbf{X}_{☾\mathcal{I}}\, \mathbf{I}^{-1}_{☾\mathcal{I}}\, \boldsymbol{h}_{☾\mathcal{I}} = \mathbf{X}_{☾\mathcal{I}}\, \mathbf{T}_{☾}\, \mathbf{I}^{-1}_{☾\mathcal{F}}\, \mathbf{T}^{T}_{☾}\, \boldsymbol{h}_{☾\mathcal{I}} \overset{\text{def}}{=} \mathbf{Y}_{☾\mathcal{I}}\, \boldsymbol{h}_{☾\mathcal{I}} \ . \end{aligned} \tag{3.91}$$

Note that eqns. (3.91) are easy to evaluate: the inverse of the inertia tensor in the body-fixed coordinate system is diagonal (the computation of the inverse is thus trivial). The transformation matrices $\mathbf{T}_{\dots}$ have to be evaluated at each integration step.

Equations (3.85), (3.88) and (3.91) represent the complete set of equations of motion of the three-body problem Earth-Moon-Sun. They describe the geocentric orbital motions of the centers of mass of the Moon and the Sun through eqn. (3.85) and the rotational motion of the Earth and Moon through eqns. (3.88) and (3.91). The complete system of equations thus consists of $2 \times 3 = 6$ second order differential equations (orbital motion), $2 \times 3 = 6$ first order equations for the angular momenta of Earth and Moon, and $2 \times 3 = 6$ first order differential equations for the Euler angles. Taking into account that each second order system may be transformed into a first order system of twice the dimension of the second order system, we have derived a system of $d = 24$ scalar, first order differential equations defining the generalized three-body problem Earth-Moon-Sun.

Note that the entire system of $d = 24$ first order equations is coupled: The geocentric position vectors of Sun and Moon are needed to evaluate the right-hand sides of the eqns. (3.88), and the orientation of the bodies is required to evaluate the right-hand sides of equations (3.85) describing the orbital motion.

In order to solve the system of equations we have to specify the initial state of the system, i.e., we have to provide the following quantities:

$$\begin{aligned} \boldsymbol{r}_{\odot}(t_0) &= \boldsymbol{r}_{\odot_0} \ , & \dot{\boldsymbol{r}}_{\odot}(t_0) &= \dot{\boldsymbol{r}}_{\odot_0} \\ \boldsymbol{r}_{☾}(t_0) &= \boldsymbol{r}_{☾_0} \ , & \dot{\boldsymbol{r}}_{☾}(t_0) &= \dot{\boldsymbol{r}}_{☾_0} \\ \boldsymbol{h}_{♁}(t_0) &= \boldsymbol{h}_{♁_0} \ , & \boldsymbol{h}_{☾}(t_0) &= \boldsymbol{h}_{☾_0} \\ \Psi_{♁}(t_0) &= \Psi_{♁_0} \ , & \Psi_{☾}(t_0) &= \Psi_{☾_0} \\ \varepsilon_{♁}(t_0) &= \varepsilon_{♁_0} \ , & \varepsilon_{☾}(t_0) &= \varepsilon_{☾_0} \\ \Theta_{♁}(t_0) &= \Theta_{♁_0} \ , & \Theta_{☾}(t_0) &= \Theta_{☾_0} \ . \end{aligned} \tag{3.92}$$

Instead of specifying the angular momentum at time t_0 it may be more convenient to specify the corresponding angular velocity vectors $\boldsymbol{\omega}_{♁}$ and $\boldsymbol{\omega}_{☾}$ and to use relation (3.73) to derive the initial state of the angular momentum vectors.

In order to use the equations of orbital and rotational motion as derived in this section, we need to know the volumes occupied by the bodies and the density distribution within these bodies. If we want to solve the initial value problem defined by the above equivalent to the 24 scalar differential equations and by the initial state of the system, we would have to evaluate the integrals on the right-hand sides of the equations of motion at each integration step. For a complicated density distribution this may be a formidable task. Also, in general, we do not have the density distributions of the celestial bodies readily available. We will therefore have to find approximations for the right-hand sides of the equations of motion which depend only on some global characteristics of the bodies, in particular on the total mass and the principal moments of inertia. Such developments will be studied in section 3.3.5.

3.3.4 The Equations of Motion in the Body-Fixed Systems

Taking into account the defining equation (3.73) for the angular momentum vector, it is easy to transform the equations (3.90) into equations for the angular velocities (and no longer for the angular momenta) for Earth and Moon. The left-hand sides of eqn. (3.90) may be transformed as follows:

$$\dot{\boldsymbol{h}}_{\mathcal{I}} = \frac{d}{dt}\left\{ \mathbf{I}_{\mathcal{I}}\,\boldsymbol{\omega}_{\mathcal{I}} \right\} = \frac{d}{dt}\left\{ \mathbf{T}\,\mathbf{I}_{\mathcal{F}}\,\mathbf{T}^T\,\boldsymbol{\omega}_{\mathcal{I}} \right\} . \tag{3.93}$$

The transformation law (3.77) for the inertia tensor was used in the last step.

One such relation stands for the Earth and one for the Moon. It is important to note in this context, that the above expression refers to the inertial system (more precisely to a geo- resp. selenocentric system, at all times parallel to the inertial system). In view of the definition (3.56) of the transformation matrix $\mathbf{T}$ we may conclude:

$$\begin{aligned} \boldsymbol{\omega}_{\mathcal{F}} &= \mathbf{T}^T\,\boldsymbol{\omega}_{\mathcal{I}} \\ \dot{\boldsymbol{h}}_{\mathcal{F}} &= \mathbf{T}^T\,\dot{\boldsymbol{h}}_{\mathcal{I}}\ . \end{aligned} \tag{3.94}$$

With these equations the relation between the first derivative (3.93) of the angular momentum and the angular velocity vector may be transferred easily into the body-fixed reference system:

$$\dot{\boldsymbol{h}}_{\mathcal{F}} = \mathbf{T}^T\,\frac{d}{dt}\left\{ \mathbf{T}\,\mathbf{I}_{\mathcal{F}}\,\boldsymbol{\omega}_{\mathcal{F}} \right\} = \mathbf{I}_{\mathcal{F}}\,\dot{\boldsymbol{\omega}}_{\mathcal{F}} \;+\; \left[\mathbf{T}^T\dot{\mathbf{T}}\right]_{\mathcal{F}}\,\mathbf{I}_{\mathcal{F}}\,\boldsymbol{\omega}_{\mathcal{F}}\ . \tag{3.95}$$

Using the fact that $\left[\mathbf{T}^T\dot{\mathbf{T}}\right]_{\mathcal{F}} = \left[\dot{\mathbf{T}}\mathbf{T}^T\right]_{\mathcal{F}}^T$ and in view of eqns. (3.69) and (3.78), the following explicit relation is established for expression (3.95) in the Earth-fixed (or Moon-fixed) system. For the first derivative of the Earth's angular momentum we obtain:

$$\dot{\boldsymbol{h}}_{\text{♁}\mathcal{F}} = \begin{pmatrix} A_{\text{♁}}\ \dot{\omega}_{\text{♁}\mathcal{F}_1} + (C_{\text{♁}} - B_{\text{♁}})\ \omega_{\text{♁}\mathcal{F}_2}\omega_{\text{♁}\mathcal{F}_3} \\ B_{\text{♁}}\ \dot{\omega}_{\text{♁}\mathcal{F}_2} + (A_{\text{♁}} - C_{\text{♁}})\ \omega_{\text{♁}\mathcal{F}_3}\omega_{\text{♁}\mathcal{F}_1} \\ C_{\text{♁}}\ \dot{\omega}_{\text{♁}\mathcal{F}_3} + (B_{\text{♁}} - A_{\text{♁}})\ \omega_{\text{♁}\mathcal{F}_1}\omega_{\text{♁}\mathcal{F}_2} \end{pmatrix} . \tag{3.96}$$

The differential equations (3.90) for the angular momentum in the inertial system thus have a simple counterpart as equations for the angular momentum in the body-fixed systems (Earth and Moon):

$$\begin{aligned} \begin{pmatrix} A_{\text{♁}}\ \dot{\omega}_{\text{♁}\mathcal{F}_1} + (C_{\text{♁}} - B_{\text{♁}})\ \omega_{\text{♁}\mathcal{F}_2}\omega_{\text{♁}\mathcal{F}_3} \\ B_{\text{♁}}\ \dot{\omega}_{\text{♁}\mathcal{F}_2} + (A_{\text{♁}} - C_{\text{♁}})\ \omega_{\text{♁}\mathcal{F}_3}\omega_{\text{♁}\mathcal{F}_1} \\ C_{\text{♁}}\ \dot{\omega}_{\text{♁}\mathcal{F}_3} + (B_{\text{♁}} - A_{\text{♁}})\ \omega_{\text{♁}\mathcal{F}_1}\omega_{\text{♁}\mathcal{F}_2} \end{pmatrix} &= \boldsymbol{\ell}_{\odot\text{♁}\mathcal{F}} + \boldsymbol{\ell}_{\text{☾}\text{♁}\mathcal{F}} \\ \begin{pmatrix} A_{\text{☾}}\ \dot{\omega}_{\text{☾}\mathcal{F}_1} + (C_{\text{☾}} - B_{\text{☾}})\ \omega_{\text{☾}\mathcal{F}_2}\omega_{\text{☾}\mathcal{F}_3} \\ B_{\text{☾}}\ \dot{\omega}_{\text{☾}\mathcal{F}_2} + (A_{\text{☾}} - C_{\text{☾}})\ \omega_{\text{☾}\mathcal{F}_3}\omega_{\text{☾}\mathcal{F}_1} \\ C_{\text{☾}}\ \dot{\omega}_{\text{☾}\mathcal{F}_3} + (B_{\text{☾}} - A_{\text{☾}})\ \omega_{\text{☾}\mathcal{F}_1}\omega_{\text{☾}\mathcal{F}_2} \end{pmatrix} &= \boldsymbol{\ell}_{\odot\text{☾}\mathcal{F}} + \boldsymbol{\ell}_{\text{♁}\text{☾}\mathcal{F}} \ . \end{aligned} \tag{3.97}$$

It should be noted that eqns. (3.97) are equations in the components referring to the resp. body-fixed systems, i.e., all the vector components have to refer to the correct body-fixed coordinate system.

Eqns. (3.97) are truly remarkable: in the absence of external torques, the components of the angular velocity vector may be established in the body-fixed system *without* the knowing the motion of the rotation axis in space. As soon as torques are present, it is of course no longer possible to establish the angular velocity independently of the Euler angles. Therefore the equations (3.97) should not be considered individually, but together with the so-called kinematical relations, which now have to be related to the body-fixed system. Eqns. (3.91) have to be replaced by the somewhat simpler relations

$$\begin{aligned} \begin{pmatrix} \dot{\Psi}_{\text{♁}} \\ \dot{\varepsilon}_{\text{♁}} \\ \dot{\Theta}_{\text{♁}} \end{pmatrix} &= \mathbf{X}_{\text{♁}\mathcal{F}}\, \mathbf{I}_{\text{♁}\mathcal{F}}^{-1}\, \boldsymbol{h}_{\text{♁}\mathcal{F}} \overset{\text{def}}{=} \mathbf{X}_{\text{♁}\mathcal{F}}\, \boldsymbol{\omega}_{\text{♁}\mathcal{F}} \\ \begin{pmatrix} \dot{\Psi}_{\text{☾}} \\ \dot{\varepsilon}_{\text{☾}} \\ \dot{\Theta}_{\text{☾}} \end{pmatrix} &= \mathbf{X}_{\text{☾}\mathcal{F}}\, \mathbf{I}_{\text{☾}\mathcal{F}}^{-1}\, \boldsymbol{h}_{\text{☾}\mathcal{F}} \overset{\text{def}}{=} \mathbf{X}_{\text{☾}\mathcal{F}}\, \boldsymbol{\omega}_{\text{☾}\mathcal{F}} \ . \end{aligned} \tag{3.98}$$

Equations (3.97) are the classical Euler equations for the rotation of a rigid planet. It is interesting to note that Euler (as it is natural) first derived his equations in the inertial system in the year 1750 in an article called *Découverte d'un nouveau principe de mécanique* ([34]). The new principle actually consisted of the insight that only the equation "change of linear momentum = sum of forces acting on a point mass" is required to understand the motion of a rigid body, because it is always possible to create any rigid body by superposition of small particles. Euler did *not* recognize at that time that the equations of motion in the inertial system expressed the fundamental law *change of angular momentum of a body = sum of the torques acting on the body.* Limited by his computational tools – a computer at that time was a human being – he did not gain too much detailed insight into the rotation of planets. This remark is perhaps not justified. After all he understood very well that the motion of the rotation axis in inertial space (i.e., precession and nutation) is explained by the equations he derived. Euler published the equations for the rotation of a rigid body in the classical form, i.e., referred to the body-fixed PAI-system, only in 1765 [36] in the famous article *Du mouvement de rotation des corps solides autour d'un axe variable* presented to the Berlin Academy on November 9, 1758 and published in 1765. It is amazing that the mathematical form of the equations of motion for point masses *and* for rigid bodies, as commonly used today, are both due to Leonhard Euler.

3.3.5 Development of the Equations of Motion

This section is rather technical. The one and only purpose consists of developing the integrals on the right-hand sides of the equations of the orbital and rotational motion and to approximate the result by simple expressions. Readers not interested in such technicalities just may inspect the final results (eqns. (3.118) for the orbital motion, eqns. (3.124) for the rotational motion). Let us point out, on the other hand, that the subsequent developments give insight into the structure of the equations.

The equations of motion, as derived in the two preceding sections, are not handy to use. Fortunately, the integrals on the right-hand sides of the equations of motion may be approximated using only the masses and the three principal moments of inertia of Earth and Moon. A typical denominator in the equations of motion has the general form

$$\frac{1}{|\boldsymbol{\Delta} - \boldsymbol{r}|} \,, \qquad \text{where} \qquad |\boldsymbol{\Delta}| = \Delta \ll r = |\boldsymbol{r}| \ . \tag{3.99}$$

This quantity may be developed into a rapidly convergent series (see, e.g., [25])

$$\frac{1}{|\boldsymbol{\Delta}-\boldsymbol{r}|}=\frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{\mathit{\Delta}}{r}\right)^{n}P_n\left(\frac{\boldsymbol{r}\cdot\boldsymbol{\Delta}}{r\mathit{\Delta}}\right)=\frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{\mathit{\Delta}}{r}\right)^{n}P_n(\cos\phi)\ , \qquad (3.100)$$

using the small quantity $\frac{\mathit{\Delta}}{r}$ as the argument of the development. ϕ is the angle between the vectors $\boldsymbol{r}$ and $\boldsymbol{\Delta}$, and $P_n(x)$ are *Legendre polynomials* (named after Adrien Marie Legendre (1752–1833)) of degree n:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \tfrac{3}{2}\,x^2-\tfrac{1}{2} \\ &\cdots\ . \end{aligned} \qquad (3.101)$$

In view of the fact that the distances between Earth, Moon and Sun are large compared to the physical sizes of Earth and Moon, we may confine ourselves to the approximation up to terms of order $n=2$ in the above series development:

$$\frac{1}{|\boldsymbol{\Delta}-\boldsymbol{r}|}=\frac{1}{r}\left\{1\ +\ \frac{\boldsymbol{r}\cdot\boldsymbol{\Delta}}{r^2}\ -\ \frac{1}{2}\,\frac{\mathit{\Delta}^2}{r^2}\ +\ \frac{3}{2}\,\frac{(\boldsymbol{r}\cdot\boldsymbol{\Delta})^2}{r^4}\right\}\ . \qquad (3.102)$$

We will also need the third power of the above quantities:

$$\frac{1}{|\boldsymbol{\Delta}-\boldsymbol{r}|^3}=\frac{1}{r^3}\left\{1\ +\ 3\,\frac{\boldsymbol{r}\cdot\boldsymbol{\Delta}}{r^2}\ -\ \frac{3}{2}\,\frac{\mathit{\Delta}^2}{r^2}\ +\ \frac{15}{2}\,\frac{(\boldsymbol{r}\cdot\boldsymbol{\Delta})^2}{r^4}\right\}\ . \qquad (3.103)$$

The integrals have to be evaluated in a well-defined coordinate system. Eventually, we may need the accelerations in the inertial system. It is much simpler, however, to perform the integration in the body-fixed systems, the terms involving integrals over $V_{♁}$ in the Earth-fixed, the terms with integrals over $V_{☾}$ in the Moon-fixed system. Afterwards, the accelerations are transformed back into the inertial system using the transformation matrices $\mathbf{T}_{♁}$ or $\mathbf{T}_{☾}$.

Let us first deal with the equations (3.85) for the orbital motions of the Moon and the Sun. The Sun is treated as a point mass in these equations, which is why the structure of the equation for the Sun seems slightly simpler than that for the Moon. Taking into account the development (3.102) we may approximate the first integral for the equation of motion for the Sun on the right-hand side as

$$
\begin{aligned}
\int\limits_{V_{\text{♁}}} \frac{\rho_{p_r}}{|\boldsymbol{r}_\odot - \boldsymbol{r}_p|}\, dV_{\text{♁}} &= \int\limits_{V_{\text{♁}}} \frac{\rho_{p_r}}{r_\odot} \left\{ 1 \;+\; \frac{\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p}{r_\odot^2} \;-\; \frac{1}{2}\,\frac{r_p^2}{r_\odot^2} \;+\; \frac{3}{2}\,\frac{(\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p)^2}{r_\odot^4} \right\} dV_{\text{♁}} \\
&= \frac{1}{r_\odot} \;-\; \frac{1}{2r_\odot^3} \int\limits_{V_{\text{♁}}} \rho_{p_r}\, r^2 \; dV_{\text{♁}} \\
&\quad + \frac{3}{2\, r_\odot^5} \int\limits_{V_{\text{♁}}} \rho_{p_r} \left\{ r_{\odot_1}^2 r_1^2 \;+\; r_{\odot_2}^2 r_2^2 \;+\; r_{\odot_3}^2 r_3^2 \right\} \; dV_{\text{♁}} \; .
\end{aligned}
\tag{3.104}
$$

The integral over the first term reduces to $\frac{1}{r_\odot}$ (due to the definition of ρ_{p_r}), the second vanishes because the body-fixed system has the origin in the center of mass. Using the definitions for the principal moments of inertia (3.79) and taking into account the relations (3.80), the above approximation assumes the form

$$
\int\limits_{V_{\text{♁}}} \frac{\rho_{p_r}}{|\boldsymbol{r}_\odot - \boldsymbol{r}_p|}\, dV_{\text{♁}} = \frac{1}{r_\odot} \;+\; \frac{A_{\text{♁}} + B_{\text{♁}} + C_{\text{♁}}}{2M\, r_\odot^3} \;-\; \frac{3\, \left(r_{\odot_1}^2 A_{\text{♁}} \;+\; r_{\odot_2}^2 B_{\text{♁}} \;+\; r_{\odot_3}^2 C_{\text{♁}} \right)}{2M\, r_\odot^5} \; .
\tag{3.105}
$$

Formula (3.105) may be brought into a standard form by

$$
I_{\boldsymbol{e}_\odot} = \boldsymbol{e}_{\odot_1}^2 A_{\text{♁}} \;+\; \boldsymbol{e}_{\odot_2}^2 B_{\text{♁}} \;+\; \boldsymbol{e}_{\odot_3}^2 C_{\text{♁}} \; .
\tag{3.106}
$$

$I_{\boldsymbol{e}_\odot}$ is the moment of inertia of the Earth in the direction $\boldsymbol{e}_\odot$, the geocentric unit vector to the Sun. Using this result in eqn. (3.105) we obtain:

$$
\int\limits_{V_{\text{♁}}} \frac{\rho_{p_r}}{|\boldsymbol{r}_\odot - \boldsymbol{r}_p|}\, dV_{\text{♁}} = \frac{1}{r_\odot} \;+\; \frac{A_{\text{♁}} + B_{\text{♁}} + C_{\text{♁}} - 3\, I_{\boldsymbol{e}_\odot}}{2M\, r_\odot^3} \; .
\tag{3.107}
$$

Because the principal moments of inertia of the Earth have very similar numerical values (see Table II- 2.1), the moment of inertia in direction $\boldsymbol{e}_\odot$ will not differ much from the three principal moments of inertia. As a consequence, $A_{\text{♁}} + B_{\text{♁}} + C_{\text{♁}} - 3\, I_{\boldsymbol{e}_\odot}$ will be a small quantity, of the order of the differences between the principal moments of inertia. Formula (3.107), after multiplication with GM, represents the potential function of the attractive force exerted by the mass distribution ($V_{\text{♁}}$ and ρ_p) on a point mass with barycentric position vector $\boldsymbol{r}_\odot$. The formula is known as *MacCullagh's formula* due to James MacCullagh (1809–1847).

Formulas (3.105) or (3.107), when written in this form, have to be evaluated in the Earth-fixed system. This circumstance has to be observed when taking the gradient of this potential function (as required in the equations of motion). Although it would be possible to develop formulae for performing these operations directly in the inertial system (see, e.g., [14]) we prefer to use the following formulation based on taking the gradient in the Earth-fixed system and then to transform the result back into the inertial system using matrix $\mathbf{T}_{\earth} \overset{\text{def}}{=} \mathbf{T}_{\earth}(\mathit{\Psi}_{\earth}, \varepsilon_{\earth}, \mathit{\Theta}_{\earth})$.

$$\begin{aligned}\nabla_{\odot} \int\limits_{V_{\earth}} \frac{\rho_{p_r}}{|\boldsymbol{r}_{\odot} - \boldsymbol{r}_p|} \, dV_{\earth} = \mathbf{T}_{\earth} \, \nabla_{\odot} \Bigg\{ \frac{1}{r_{\odot}} &+ \frac{A_{\earth} + B_{\earth} + C_{\earth}}{2M \, r_{\odot}^3} \\ &- \frac{3 \left(r_{\odot_1}^2 A_{\earth} + r_{\odot_2}^2 B_{\earth} + r_{\odot_3}^2 C_{\earth} \right)}{2M \, r_{\odot}^5} \Bigg\} \, .\end{aligned} \tag{3.108}$$

All component matrices and the gradient on the left-hand side refer to the inertial, all symbols on the right-hand side to the Earth-fixed system.

The result (3.108) may be transcribed easily to the second term in the equation of motion for the Sun in eqns. (3.85): the geocentric position vector $\boldsymbol{r}_{\odot}$ of the Sun has to be replaced by the selenocentric position vector $\boldsymbol{r}_{\leftmoon\odot} \overset{\text{def}}{=} \boldsymbol{r}_{\odot} - \boldsymbol{r}_{\leftmoon}$, and the Earth-related quantities (mass, density, volume) have to be replaced by the corresponding Moon-related quantities:

$$\begin{aligned}\nabla_{\odot} \int\limits_{V_{\leftmoon}} \frac{\rho_{\wp_r}}{|\boldsymbol{r}_{\leftmoon\odot} - \boldsymbol{r}_{\leftmoon\wp}|} \, dV_{\leftmoon} = \mathbf{T}_{\leftmoon} \, \nabla_{\odot} \Bigg\{ \frac{1}{r_{\leftmoon\odot}} &+ \frac{A_{\leftmoon} + B_{\leftmoon} + C_{\leftmoon}}{2m_{\leftmoon} \, r_{\leftmoon\odot}^3} \\ &- \frac{3 \left(r_{\leftmoon\odot_1}^2 A_{\earth} + r_{\leftmoon\odot_2}^2 B_{\earth} + r_{\leftmoon\odot_3}^2 C_{\earth} \right)}{2m_{\leftmoon} \, r_{\leftmoon\odot}^5} \Bigg\} \, .\end{aligned} \tag{3.109}$$

The last term to be considered in the equations of motion for the Sun (3.85) contains a double integration over the volumes occupied by Earth and Moon. First, the potential function of this term is considered. Equation (3.103) has to be used to approximate the denominator on the left-hand side. In an attempt to reduce the formalism we put

$$\boldsymbol{\mathit{\Delta}} \overset{\text{def}}{=} \boldsymbol{r}_p - \boldsymbol{r}_{\leftmoon\wp} \tag{3.110}$$

and obtain (observe the sign to be chosen to be consistent with formula (3.103)):

$$\int\limits_{V_{\leftmoon}}\int\limits_{V_{\earth}} \rho_{\wp_r}\rho_{p_r}\, \frac{(\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p)\cdot\boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p|^3}\; dV_{\earth}\, dV_{\leftmoon} = \int\limits_{V_{\leftmoon}}\int\limits_{V_{\earth}} \frac{\rho_{\wp_r}\rho_{p_r}}{r_{\leftmoon}^3} \left\{ 1 + \frac{3}{r_{\leftmoon}^2}\,\boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{\varDelta} \right. \tag{3.111}$$
$$\left. - \frac{3}{2\,r_{\leftmoon}^2}\,\varDelta^2 + \frac{15}{2\,r_{\leftmoon}^4}\,(\boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{\varDelta})^2 \right\} (\boldsymbol{r}_{\leftmoon}-\boldsymbol{\varDelta})\cdot\boldsymbol{r}_{\odot}\; dV_{\earth}\, dV_{\leftmoon}\ .$$

All terms in the brackets $\{\ldots\}$, except for the second, give non-zero contributions when multiplied with $\boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{r}_{\odot}$, only the second term contributes (considering only terms up to the second order in $\varDelta$) when multiplied with $\boldsymbol{\varDelta}\cdot\boldsymbol{r}_{\odot}$. Assuming again that we evaluate the integrals either in the Earth-fixed or the Moon-fixed systems and considering that all terms linear in the components of either $\boldsymbol{r}_p$ or $\boldsymbol{r}_{\leftmoon\wp}$ and the mixed terms $r_{pi}\,r_{pk}$ will give no contribution after integration, the following terms actually have to be considered of the above expressions (all other terms are marked with ... in the following equations):

$$\begin{aligned}
\varDelta^2 \quad &= r_p^2 + r_{\leftmoon\wp}^2 - \ldots \\
(\boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{\varDelta})^2 \quad &= r_{p1}^2 r_{\leftmoon_1}^2 + r_{p2}^2 r_{\leftmoon_2}^2 + r_{p3}^2 r_{\leftmoon_3}^2 + \ldots \\
&\quad + r_{\leftmoon\wp_1}^2 r_{\leftmoon_1}^2 + r_{\leftmoon\wp_2}^2 r_{\leftmoon_2}^2 + r_{\leftmoon\wp_3}^2 r_{\leftmoon_3}^2 + \ldots \\
(\boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{\varDelta})\,(\boldsymbol{r}_{\odot}\cdot\boldsymbol{\varDelta}) &= r_{p_1}^2 r_{\leftmoon_1} r_{\odot_1} + r_{p_2}^2 r_{\leftmoon_2} r_{\odot_2} + r_{p_3}^2 r_{\leftmoon_3} r_{\odot_3} + \ldots \\
&\quad + r_{\leftmoon\wp_1}^2 r_{\leftmoon_1} r_{\odot_1} + r_{\leftmoon\wp_2}^2 r_{\leftmoon_2} r_{\odot_2} + r_{\leftmoon\wp_3}^2 r_{\leftmoon_3} r_{\odot_3} + \ldots\ .
\end{aligned} \tag{3.112}$$

The relevant terms related to Earth and Moon are very nicely separated. Note that the components of the vectors $\boldsymbol{r}_{\leftmoon}$ and $\boldsymbol{r}_{\odot}$ are referred to the geocentric PAI-system, to the corresponding selenocentric PAI-system for Moon-related parts. Using the expressions (3.112) we may give eqns. (3.111) the following explicit form:

$$\int\limits_{V_{\leftmoon}}\int\limits_{V_{\earth}} \rho_{\wp_r}\,\rho_{p_r}\, \frac{(\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p)\cdot\boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{\leftmoon}+\boldsymbol{r}_{\leftmoon\wp}-\boldsymbol{r}_p|^3}\; dV_{\earth}\, dV_{\leftmoon} = \frac{1}{r_{\leftmoon}^3}\,\boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{r}_{\odot}$$
$$+ \left\{ \frac{3}{2\,r_{\leftmoon}^5\,M}\,(A_{\earth}+B_{\earth}+C_{\earth}) - \frac{15}{2\,r_{\leftmoon}^7\,M}\left(r_{\leftmoon_1}^2 A_{\earth} + r_{\leftmoon_2}^2 B_{\earth} + r_{\leftmoon_3}^2 C_{\earth}\right) \right\} \boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{r}_{\odot}$$
$$- \frac{3}{r_{\leftmoon}^5\,M}\,\{\, A_{\earth} r_{\leftmoon_1} r_{\odot_1} + B_{\earth} r_{\leftmoon_2} r_{\odot_2} + C_{\earth} r_{\leftmoon_3} r_{\odot_3} \,\}$$
$$+ \left\{ \frac{3}{2\,r_{\leftmoon}^5\,m}\,(A_{\leftmoon}+B_{\leftmoon}+C_{\leftmoon}) - \frac{15}{2\,r_{\leftmoon}^7\,m}\left(r_{\leftmoon_1}^2 A_{\leftmoon} + r_{\leftmoon_2}^2 B_{\leftmoon} + r_{\leftmoon_3}^2 C_{\leftmoon}\right) \right\} \boldsymbol{r}_{\leftmoon}\cdot\boldsymbol{r}_{\odot}$$
$$- \frac{3}{r_{\leftmoon}^5\,m}\,\{\, A_{\leftmoon} r_{\leftmoon_1} r_{\odot_1} + B_{\leftmoon} r_{\leftmoon_2} r_{\odot_2} + C_{\leftmoon} r_{\leftmoon_3} r_{\odot_3} \,\}\ . \tag{3.113}$$

Note that for $A_{♁} = B_{♁} = C_{♁}$ and $A_{☾} = B_{☾} = C_{☾}$ the above scalar function is reduced to the first term on the right-hand side, which is, as a matter of fact, the term of the classical three-body problem. Using matrix notation, the gradient of eqn. (3.113) may be written in the following elegant way:

$$\begin{aligned} &\nabla_{\odot} \int\limits_{V_{☾}} \int\limits_{V_{♁}} \rho_{\wp_r}\, \rho_{p_r}\, \frac{(\boldsymbol{r}_{☾} + \boldsymbol{r}_{☾\wp} - \boldsymbol{r}_p) \cdot \boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{☾} + \boldsymbol{r}_{☾\wp} - \boldsymbol{r}_p|^3}\, dV_{♁}\, dV_{☾} = \frac{1}{r_{☾}^3}\, \boldsymbol{r}_{☾} \\ &+ \frac{3}{2\, r_{☾}^5\, M}\, \mathbf{T}_{♁} \left\{ \left[\left(A_{♁} + B_{♁} + C_{♁} - \frac{5}{r_{☾}^2}\, \boldsymbol{r}_{☾}^T\, \mathbf{I}_{♁\mathcal{F}}\, \boldsymbol{r}_{☾} \right) \mathbf{E} \;+\; 2\, \mathbf{I}_{♁\mathcal{F}} \right] \boldsymbol{r}_{☾} \right\} \\ &+ \frac{3}{2\, r_{☾}^5\, m}\, \mathbf{T}_{☾} \left\{ \left[\left(A_{☾} + B_{☾} + C_{☾} - \frac{5}{r_{☾}^2}\, \boldsymbol{r}_{☾}^T\, \mathbf{I}_{☾\mathcal{F}}\, \boldsymbol{r}_{☾} \right) \mathbf{E} \;+\; 2\, \mathbf{I}_{☾\mathcal{F}} \right] \boldsymbol{r}_{☾} \right\} . \end{aligned} \tag{3.114}$$

Equation (3.114) refers to the inertial system. The first bracket on the right-hand side has to be evaluated in the geocentric, Earth-fixed PAI-system, the second bracket in the corresponding selenocentric system.

With eqns. (3.108), (3.109) and (3.114) we may approximate the equations of motion for the Sun in the three-body problem using the masses and the principal moments of inertia. Using the work which was necessary to perform this task it is now comparatively easy to find the corresponding approximation for the motion of the Moon in eqns. (3.85). Let us first compute the two-body term of this motion:

$$\begin{aligned} &\nabla_{☾} \int\limits_{V_{☾}} \int\limits_{V_{♁}} \frac{\rho_{p_r}\, \rho_{\wp_r}}{|\boldsymbol{r}_{☾} + \boldsymbol{r}_{☾\wp} - \boldsymbol{r}_p|}\, dV_{♁}\, dV_{☾} = \nabla_{☾} \left\{ \frac{1}{r_{☾}} \right\} \\ &+ \mathbf{T}_{♁}\, \nabla_{☾} \left\{ \frac{A_{♁} + B_{♁} + C_{♁}}{2\, M\, r_{☾}^3} - \frac{3\, \left(r_{☾_1}^2 A_{♁} + r_{☾_2}^2 B_{♁} + r_{☾_3}^2 C_{♁} \right)}{2M\, r_{☾}^5} \right\} \\ &+ \mathbf{T}_{☾}\, \nabla_{☾} \left\{ \frac{A_{☾} + B_{☾} + C_{☾}}{2\, m_{☾}\, r_{☾}^3} - \frac{3\, \left(r_{☾_1}^2 A_{☾} + r_{☾_2}^2 B_{☾} + r_{☾_3}^2 C_{☾} \right)}{2\, m_{☾}\, r_{☾}^5} \right\} . \end{aligned} \tag{3.115}$$

The first bracket on the right-hand side may be directly evaluated in the inertial system, the second refers to the geocentric, the last to the selenocentric PAI-systems.

Equation (3.115), after multiplication with $G(M + m_{☾})$, contains the potential function for the generalized two-body problem Earth-Moon. The relative motion of the centers of mass of two bodies of arbitrary shape is thus described by a potential function which is the "weighted" sum of the potential functions as given by MacCullagh's formulas for the Earth and the Moon individually, the weights being $\frac{G(M+m_{\leftmoon})}{M}$ for the term corresponding to the Earth, $\frac{G(M+m_{\leftmoon})}{m_{\leftmoon}}$ for the term corresponding to the Moon. This formula is, e.g., reported by Brouwer and Clemence [27].

The first part of the perturbation acceleration in the equations of motion for the Moon is in essence already given by formula (3.109):

$$\nabla_{\leftmoon} \int\limits_{V_{\leftmoon}} \frac{\rho_{\wp_r}}{|\boldsymbol{r}_{\leftmoon\odot} - \boldsymbol{r}_{\leftmoon\wp}|} \, dV_{\leftmoon} =$$

$$\mathbf{T}_{\leftmoon} \, \nabla_{\leftmoon} \left\{ \frac{1}{r_{\leftmoon\odot}} + \frac{A_{\leftmoon} + B_{\leftmoon} + C_{\leftmoon}}{2\, m_{\leftmoon} \, r_{\leftmoon\odot}^3} - \frac{3\, \left(r_{\leftmoon\odot_1}^2 A_{\earth} + r_{\leftmoon\odot_2}^2 B_{\earth} + r_{\leftmoon\odot_3}^2 C_{\earth} \right)}{2\, m_{\leftmoon} \, r_{\leftmoon\odot}^5} \right\} , \tag{3.116}$$

and the indirect part of the perturbative acceleration may be transcribed from eqn. (3.114):

$$\nabla_{\leftmoon} \int\limits_{V_{\earth}} \rho_{p_r} \frac{(\boldsymbol{r}_{\odot} - \boldsymbol{r}_p) \cdot \boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{\odot} - \boldsymbol{r}_p|^3} \, dV_{\earth} = \frac{1}{r_{\odot}^3} \, \boldsymbol{r}_{\odot}$$

$$+ \frac{3}{2 r_{\odot}^5 \, M} \, \mathbf{T}_{\earth} \left\{ \left[\left(A_{\earth} + B_{\earth} + C_{\earth} - \frac{5}{r_{\odot}^2} \, \boldsymbol{r}_{\odot}^T \, \mathbf{I}_{\earth_{\mathcal{F}}} \, \boldsymbol{r}_{\odot} \right) \mathbf{E} \; + \; 2\, \mathbf{I}_{\earth_{\mathcal{F}}} \right] \boldsymbol{r}_{\odot} \right\} . \tag{3.117}$$

With this result the equations of motion for the generalized three-body problem may be summarized in the approximation sought in this section. The terms below are arranged to let the point mass approximation precede the terms proportional to the principal moments of inertia. The latter terms are expressed as gradients of brackets $\{\ldots\}$. The gradients have to be evaluated in the corresponding body-fixed PAI-system. Afterwards the term is transformed into the inertial system using the matrices $\mathbf{T}_{\earth}$ and $\mathbf{T}_{\leftmoon}$. When solving the equations of motion with numerical techniques it is thus necessary to transform the geocentric position vectors of Sun and Moon at each integration step into the geocentric and the selenocentric PAI-system, to evaluate the gradients in these systems, and to transform the result back into the inertial system.

$$
\begin{aligned}
\ddot{\boldsymbol{r}}_{\leftmoon} =& -G\,(M+m_{\leftmoon})\,\frac{\boldsymbol{r}_{\leftmoon}}{r_{\leftmoon}^3} - G\,m_{\odot}\left[\frac{\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\odot}}{|\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\odot}|^3}+\frac{\boldsymbol{r}_{\odot}}{r_{\odot}^3}\right]\\
&+\frac{G\,(M+m_{\leftmoon})}{2\,M}\mathbf{T}_{\earth}\nabla_{\leftmoon}\left\{\frac{A_{\earth}+B_{\earth}+C_{\earth}}{r_{\leftmoon}^3}-\frac{3\,\left(r_{\leftmoon_1}^2A_{\earth}+r_{\leftmoon_2}^2B_{\earth}+r_{\leftmoon_3}^2C_{\earth}\right)}{r_{\leftmoon}^5}\right\}\\
&+\frac{G\,(M+m_{\leftmoon})}{2m_{\leftmoon}}\mathbf{T}_{\leftmoon}\nabla_{\leftmoon}\left\{\frac{A_{\leftmoon}+B_{\leftmoon}+C_{\leftmoon}}{r_{\leftmoon}^3}-\frac{3\,\left(r_{\leftmoon_1}^2A_{\leftmoon}+r_{\leftmoon_2}^2B_{\leftmoon}+r_{\leftmoon_3}^2C_{\leftmoon}\right)}{r_{\leftmoon}^5}\right\}\\
&+\frac{Gm_{\odot}}{2\,m_{\leftmoon}}\mathbf{T}_{\leftmoon}\nabla_{\leftmoon\odot}\left\{\frac{A_{\leftmoon}+B_{\leftmoon}+C_{\leftmoon}}{r_{\leftmoon\odot}^3}-\frac{3\,\left(r_{\leftmoon\odot_1}^2A_{\earth}+r_{\leftmoon\odot_2}^2B_{\earth}+r_{\leftmoon\odot_3}^2C_{\earth}\right)}{r_{\leftmoon\odot}^5}\right\}\\
&-\frac{3\,G\,m_{\odot}}{2\,M\,r_{\odot}^5}\,\mathbf{T}_{\earth}\left\{\left[\left(A_{\earth}+B_{\earth}+C_{\earth}-\frac{5}{r_{\odot}^2}\,\boldsymbol{r}_{\odot}^T\,\mathbf{I}_{\earth_{\mathcal{F}}}\,\boldsymbol{r}_{\odot}\right)\,\mathbf{E}\,+\,2\,\mathbf{I}_{\earth_{\mathcal{F}}}\right]\boldsymbol{r}_{\odot}\right\}\\
\ddot{\boldsymbol{r}}_{\odot} =& -G\,(M+m_{\odot})\,\frac{\boldsymbol{r}_{\odot}}{r_{\odot}^3} - G\,m_{\odot}\left[\frac{\boldsymbol{r}_{\odot}-\boldsymbol{r}_{\leftmoon}}{|\boldsymbol{r}_{\odot}-\boldsymbol{r}_{\leftmoon}|^3}+\frac{\boldsymbol{r}_{\leftmoon}}{r_{\leftmoon}^3}\right]\\
&+\frac{G(M+m_{\odot})}{2\,M}\mathbf{T}_{\earth}\nabla_{\odot}\left\{\frac{A_{\earth}+B_{\earth}+C_{\earth}}{r_{\odot}^3}-\frac{3\,\left(r_{\odot_1}^2A_{\earth}+r_{\odot_2}^2B_{\earth}+r_{\odot_3}^2C_{\earth}\right)}{r_{\odot}^5}\right\}\\
&-\frac{3\,G\,m_{\leftmoon}}{2\,M\,r_{\leftmoon}^5}\,\mathbf{T}_{\earth}\left\{\left[\left(A_{\earth}+B_{\earth}+C_{\earth}-\frac{5}{r_{\leftmoon}^2}\,\boldsymbol{r}_{\leftmoon}^T\,\mathbf{I}_{\earth_{\mathcal{F}}}\,\boldsymbol{r}_{\leftmoon}\right)\,\mathbf{E}\,+\,2\,\mathbf{I}_{\earth_{\mathcal{F}}}\right]\,\boldsymbol{r}_{\leftmoon}\right\}\\
&-\frac{3\,G\,m_{\leftmoon}}{2\,m_{\leftmoon}\,r_{\leftmoon}^5}\,\mathbf{T}_{\leftmoon}\left\{\left[\left(A_{\leftmoon}+B_{\leftmoon}+C_{\leftmoon}-\frac{5}{r_{\leftmoon}^2}\,\boldsymbol{r}_{\leftmoon}^T\,\mathbf{I}_{\leftmoon_{\mathcal{F}}}\,\boldsymbol{r}_{\leftmoon}\right)\,\mathbf{E}\,+\,2\,\mathbf{I}_{\leftmoon_{\mathcal{F}}}\right]\,\boldsymbol{r}_{\leftmoon}\right\}\;.
\end{aligned}
\tag{3.118}
$$

The generalized equations of motion (3.118) for the geocentric motion of the Sun and Moon are well structured: If only the first line in each of the equations is taken into account, we obtain the equations of motion for the classical three-body problem with point masses. If only the first three lines of the equation for the Moon are taken into account, the equations for the generalized two-body problem Earth-Moon are obtained.

What still remains to be done is the derivation of the equations (3.88) for the rotation of Earth and Moon in the approximation used above for the orbital motion. As a matter of fact, we only have to deal with the right-hand sides of these equations, i.e., with the torques. In order to make our derivations as simple as possible, we compute the torques in the corresponding PAI-systems. Using the same development as in eqn. (3.104) we may write the first term in (3.88) as

$$\begin{aligned}&\int\limits_{V_{\text{♁}}} \rho_p \, \nabla_\odot \left\{ \frac{1}{|\boldsymbol{r}_\odot - \boldsymbol{r}_p|} \right\} \times \boldsymbol{r}_\odot \; dV_{\text{♁}} = \\ &\int\limits_{V_{\text{♁}}} \rho_p \, \nabla_\odot \left\{ \frac{1}{r_\odot} + \frac{\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p}{r_\odot^3} - \frac{1}{2}\,\frac{r_p^2}{r_\odot^3} + \frac{3}{2}\,\frac{(\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p)^2}{r_\odot^5} \right\} \times \boldsymbol{r}_\odot \; dV_{\text{♁}} \; .\end{aligned} \tag{3.119}$$

What at first sight looks like a formidable task becomes rather simple, because $\nabla(r^n) \times \boldsymbol{r} = \boldsymbol{0}$. Therefore, we only have to consider the terms which result from taking the gradient of the numerators in the brackets $\{\ldots\}$ above. This in turn immediately implies that only the contributions due to the second and the fourth term have to be considered, because the other two do not depend on $\boldsymbol{r}_\odot$. Taking into account that the second term is linear in the components of $\boldsymbol{r}_p$, the integral related to this term will be zero (center of mass condition). Therefore, only the contribution due to the last term, considering only the dependence on $\boldsymbol{r}_\odot$ of the numerator, will be different from zero:

$$\begin{aligned}&\int\limits_{V_{\text{♁}}} \rho_p \, \nabla_\odot \left\{ \frac{1}{r_\odot} + \frac{\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p}{r_\odot^3} - \frac{1}{2}\,\frac{r_p^2}{r_\odot^3} + \frac{3}{2}\,\frac{(\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p)^2}{r_\odot^5} \right\} \times \boldsymbol{r}_\odot \; dV_{\text{♁}} \\ &= \int\limits_{V_{\text{♁}}} \frac{3\,\rho_p}{2\,r_\odot^5} \, \nabla_\odot \left\{ (\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p)^2 \right\} \times \boldsymbol{r}_\odot \; dV_{\text{♁}} \; = \int\limits_{V_{\text{♁}}} \frac{3\,\rho_p}{r_\odot^5} (\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p)\, \boldsymbol{r}_p \times \boldsymbol{r}_\odot \; dV_{\text{♁}} \\ &= \int\limits_{V_{\text{♁}}} \frac{3\,\rho_p}{r_\odot^5} \, (\boldsymbol{r}_\odot \cdot \boldsymbol{r}_p) \begin{pmatrix} r_{p2} r_{\odot 3} - r_{p3} r_{\odot 2} \\ r_{p3} r_{\odot 1} - r_{p1} r_{\odot 3} \\ r_{p1} r_{\odot 2} - r_{p2} r_{\odot 1} \end{pmatrix} dV_{\text{♁}} \; = \; \frac{3}{r_\odot^5} \begin{pmatrix} (C_{\text{♁}} - B_{\text{♁}}) r_{\odot 2} r_{\odot 3} \\ (A_{\text{♁}} - C_{\text{♁}}) r_{\odot 3} r_{\odot 1} \\ (B_{\text{♁}} - A_{\text{♁}}) r_{\odot 1} r_{\odot 2} \end{pmatrix} .\end{aligned} \tag{3.120}$$

In the last step we made use of the fact that only the diagonal terms of the inertia tensor are different from zero in the PAI-system. For the computation of the diagonal terms the formulae (3.80) were used.

The result (3.120) may be transcribed to the Sun- and Earth-induced torque on the Moon in eqns. (3.88):

$$\int\limits_{V_{\leftmoon}} \rho_{\wp} \, \nabla_\odot \left\{ \frac{1}{|\boldsymbol{r}_\odot - \boldsymbol{r}_{\leftmoon} - \boldsymbol{r}_{\leftmoon\wp}|} \right\} \times \boldsymbol{r}_\odot \; dV_{\leftmoon} = \frac{3}{r_{\odot\leftmoon}^5} \begin{pmatrix} (C_{\leftmoon} - B_{\leftmoon})\, r_{\odot\leftmoon_2}\, r_{\odot\leftmoon_3} \\ (A_{\leftmoon} - C_{\leftmoon})\, r_{\odot\leftmoon_3}\, r_{\odot\leftmoon_1} \\ (B_{\leftmoon} - A_{\leftmoon})\, r_{\odot\leftmoon_1}\, r_{\odot\leftmoon_2} \end{pmatrix} . \tag{3.121}$$

Note that the result has this form only if the Moon's PAI-system is used.

The computation of the torque exerted by the Earth on the Moon follows the same pattern as above. It is interesting to note that to the level of approximation of this section the result is the same as if the attracting mass were a point mass (and not an extended body). The result is given by

$$
\begin{aligned}
&\int\limits_{V_{\earth}}\int\limits_{V_{\leftmoon}} \rho_p\,\rho_{\wp}\,\nabla_{\leftmoon}\left\{\frac{1}{|\boldsymbol{r}_p-\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\leftmoon\wp}|}\right\}\times\left(\boldsymbol{r}_{\leftmoon\wp}+\boldsymbol{r}_{\leftmoon}\right)\;dV_{\leftmoon}\;dV_{\earth}\\
&=\;\frac{3\,m}{r^5_{\odot\leftmoon}}\begin{pmatrix}(C_{\earth}-B_{\earth})\,r_{\leftmoon_2}\,r_{\leftmoon_3}\\(A_{\earth}-C_{\earth})\,r_{\leftmoon_3}\,r_{\leftmoon_1}\\(B_{\earth}-A_{\earth})\,r_{\leftmoon_1}\,r_{\leftmoon_2}\end{pmatrix}\\
&\int\limits_{V_{\earth}}\int\limits_{V_{\leftmoon}} \rho_p\,\rho_{\wp}\nabla_{\leftmoon}\left\{\frac{1}{|\boldsymbol{r}_p-\boldsymbol{r}_{\leftmoon}-\boldsymbol{r}_{\wp}|}\right\}\times\left(\boldsymbol{r}_p-\boldsymbol{r}_{\leftmoon}\right)\;dV_{\leftmoon}\;dV_{\earth}\\
&=\;\frac{3\,M}{r^5_{\odot\leftmoon}}\begin{pmatrix}(C_{\leftmoon}-B_{\leftmoon})\,r_{\leftmoon_2}\,r_{\leftmoon_3}\\(A_{\leftmoon}-C_{\leftmoon})\,r_{\leftmoon_3}\,r_{\leftmoon_1}\\(B_{\leftmoon}-A_{\leftmoon})\,r_{\leftmoon_1}\,r_{\leftmoon_2}\end{pmatrix}\;.
\end{aligned}
\tag{3.122}
$$

With the notation

$$
\begin{aligned}
\gamma_{\earth_1}&\overset{\text{def}}{=}\frac{C_{\earth}-B_{\earth}}{A_{\earth}}\;;\quad &\gamma_{\leftmoon_1}&\overset{\text{def}}{=}\frac{C_{\leftmoon}-B_{\leftmoon}}{A_{\leftmoon}}\\
\gamma_{\earth_2}&\overset{\text{def}}{=}\frac{A_{\earth}-C_{\earth}}{B_{\earth}}\;;\quad &\gamma_{\leftmoon_2}&\overset{\text{def}}{=}\frac{A_{\leftmoon}-C_{\leftmoon}}{B_{\leftmoon}}\\
\gamma_{\earth_3}&\overset{\text{def}}{=}\frac{B_{\earth}-A_{\earth}}{C_{\earth}}\;;\quad &\gamma_{\leftmoon_3}&\overset{\text{def}}{=}\frac{B_{\leftmoon}-A_{\leftmoon}}{C_{\leftmoon}}
\end{aligned}
\tag{3.123}
$$

the equations for the rotation of Earth and Moon may be written in the resp. PAI-systems as

$$
\begin{aligned}
\begin{pmatrix}\dot\omega_{\earth_1}\\\dot\omega_{\earth_2}\\\dot\omega_{\earth_3}\end{pmatrix}+\begin{pmatrix}\gamma_{\earth_1}\,\omega_{\earth_2}\,\omega_{\earth_3}\\\gamma_{\earth_2}\,\omega_{\earth_3}\,\omega_{\earth_1}\\\gamma_{\earth_3}\,\omega_{\earth_1}\,\omega_{\earth_2}\end{pmatrix}&=+\,\frac{3\,G\,m_{\leftmoon}}{r^5_{\leftmoon}}\begin{pmatrix}\gamma_{\earth_1}\,r_{\leftmoon_2}\,r_{\leftmoon_3}\\\gamma_{\earth_2}\,r_{\leftmoon_3}\,r_{\leftmoon_1}\\\gamma_{\earth_3}\,r_{\leftmoon_1}\,r_{\leftmoon_2}\end{pmatrix}\\
&\quad+\,\frac{3\,G\,m_{\odot}}{r^5_{\odot}}\begin{pmatrix}\gamma_{\earth_1}\,r_{\odot_2}\,r_{\odot_3}\\\gamma_{\earth_2}\,r_{\odot_3}\,r_{\odot_1}\\\gamma_{\earth_3}\,r_{\odot_1}\,r_{\odot_2}\end{pmatrix}\\
\begin{pmatrix}\dot\omega_{\leftmoon_1}\\\dot\omega_{\leftmoon_2}\\\dot\omega_{\leftmoon_3}\end{pmatrix}+\begin{pmatrix}\gamma_{\leftmoon_1}\,\omega_{\leftmoon_2}\,\omega_{\leftmoon_3}\\\gamma_{\leftmoon_2}\,\omega_{\leftmoon_3}\,\omega_{\leftmoon_1}\\\gamma_{\leftmoon_3}\,\omega_{\leftmoon_1}\,\omega_{\leftmoon_2}\end{pmatrix}&=+\,\frac{3\,G\,M}{r^5_{\leftmoon}}\begin{pmatrix}\gamma_{\leftmoon_1}\,r_{\leftmoon\earth_2}\,r_{\leftmoon\earth_3}\\\gamma_{\leftmoon_2}\,r_{\leftmoon\earth_3}\,r_{\leftmoon\earth_1}\\\gamma_{\leftmoon_3}\,r_{\leftmoon\earth_1}\,r_{\leftmoon\earth_2}\end{pmatrix}\\
&\quad+\,\frac{3\,G\,m_{\odot}}{r^5_{\odot}}\begin{pmatrix}\gamma_{\leftmoon_1}\,r_{\leftmoon\odot_2}\,r_{\leftmoon\odot_3}\\\gamma_{\leftmoon_2}\,r_{\leftmoon\odot_3}\,r_{\leftmoon\odot_1}\\\gamma_{\leftmoon_3}\,r_{\leftmoon\odot_1}\,r_{\leftmoon\odot_2}\end{pmatrix}\;.
\end{aligned}
\tag{3.124}
$$

Obviously the equations show a particularly simple structure. For the solution of the equations in the inertial system we obtain

$$
\begin{aligned}
\dot{\boldsymbol{h}}_{\text{♁}} &= + \frac{3\,G\,m_{\text{☾}}}{r_{\text{☾}}^{5}}\,\mathbf{T}_{\text{♁}} \begin{pmatrix} (C_{\text{♁}} - B_{\text{♁}})\,r_{\text{☾}_2}\,r_{\text{☾}_3} \\ (A_{\text{♁}} - C_{\text{♁}})\,r_{\text{☾}_3}\,r_{\text{☾}_1} \\ (B_{\text{♁}} - A_{\text{♁}})\,r_{\text{☾}_1}\,r_{\text{☾}_2} \end{pmatrix} \\
&\quad + \frac{3\,G\,m_{\odot}}{r_{\odot}^{5}}\,\mathbf{T}_{\text{♁}} \begin{pmatrix} (C_{\text{♁}} - B_{\text{♁}})\,r_{\odot_2}\,r_{\odot_3} \\ (A_{\text{♁}} - C_{\text{♁}})\,r_{\odot_3}\,r_{\odot_1} \\ (B_{\text{♁}} - A_{\text{♁}})\,r_{\odot_1}\,r_{\odot_2} \end{pmatrix} \\
\dot{\boldsymbol{h}}_{\text{☾}} &= + \frac{3\,G\,M}{r_{\text{☾}}^{5}}\,\mathbf{T}_{\text{☾}} \begin{pmatrix} (C_{\text{☾}} - B_{\text{☾}})\,r_{\text{☾}_2}\,r_{\text{☾}_3} \\ (A_{\text{☾}} - C_{\text{☾}})\,r_{\text{☾}_3}\,r_{\text{☾}_1} \\ (B_{\text{☾}} - A_{\text{☾}})\,r_{\text{☾}_1}\,r_{\text{☾}_2} \end{pmatrix} \\
&\quad + \frac{3\,G\,m_{\odot}}{r_{\odot\text{☾}}^{5}}\,\mathbf{T}_{\text{☾}} \begin{pmatrix} (C_{\text{☾}} - B_{\text{☾}})\,r_{\text{☾}\odot_2}\,r_{\text{☾}\odot_3} \\ (A_{\text{☾}} - C_{\text{☾}})\,r_{\text{☾}\odot_3}\,r_{\text{☾}\odot_1} \\ (B_{\text{☾}} - A_{\text{☾}})\,r_{\text{☾}\odot_1}\,r_{\text{☾}\odot_2} \end{pmatrix} .
\end{aligned}
\tag{3.125}
$$

As usual the terms $\mathbf{T}_{\text{♁}}\,(\ldots)$ have to be evaluated in the Earth's PAI-system, the terms $\mathbf{T}_{\text{☾}}\,(\ldots)$ in the Moon's PAI-system.

In the above approximation the generalized three-body problem Earth-Sun-Moon is described by the equations (3.118) and either eqns. (3.124, 3.68) if the rotational motion is described in the PAI-systems or eqns. (3.125, 3.66), if this motion is described in the inertial system. These systems of equations represent the most general three-body problem with rigid bodies considered here.

3.3.6 Second Order Differential Equations for the Euler Angles $\boldsymbol{\Psi}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\Theta}$

The rotational motion of the Earth and the Moon are defined by the equations (3.88) and the corresponding kinematic Euler equations (3.68) (one set of kinematic equations must be used for Earth rotation, one for the rotation of the Moon). Together, the equations (3.88) and (3.68) form one set of $2 \cdot 6$ first-order differential equations.

When solving the generalized three-body problem, i.e., when solving simultaneously the equations for the orbital and rotational motion of Earth, Moon, and Sun, it would be preferable to transform the first order differential equation system for the Euler angles and the components of the angular velocity vector into one second order system for the three Euler angles. This can be achieved easily by taking the time derivative of Euler's kinematic equations (3.68):

$$
\begin{aligned}
\begin{pmatrix} \ddot{\Psi}_{\text{♁}} \\ \ddot{\varepsilon}_{\text{♁}} \\ \ddot{\Theta}_{\text{♁}} \end{pmatrix} &= \begin{pmatrix} -\sin\Theta_{\text{♁}}\csc\varepsilon_{\text{♁}} & -\cos\Theta_{\text{♁}}\csc\varepsilon_{\text{♁}} & 0 \\ -\cos\Theta_{\text{♁}} & +\sin\Theta_{\text{♁}} & 0 \\ \sin\Theta_{\text{♁}}\cot\varepsilon_{\text{♁}} & +\cos\Theta_{\text{♁}}\cot\varepsilon_{\text{♁}} & 1 \end{pmatrix} \dot{\boldsymbol{\omega}}_{\text{♁}\mathcal{F}}(t) \\
&+ \frac{d}{dt}\begin{pmatrix} -\sin\Theta_{\text{♁}}\csc\varepsilon_{\text{♁}} & -\cos\Theta_{\text{♁}}\csc\varepsilon_{\text{♁}} & 0 \\ -\cos\Theta_{\text{♁}} & +\sin\Theta_{\text{♁}} & 0 \\ \sin\Theta_{\text{♁}}\cot\varepsilon_{\text{♁}} & +\cos\Theta_{\text{♁}}\cot\varepsilon_{\text{♁}} & 1 \end{pmatrix} \boldsymbol{\omega}_{\text{♁}\mathcal{F}}(t) \; .
\end{aligned}
\tag{3.126}
$$

Using the differential equations (3.88) for the components of the angular velocity vectors and Euler's kinematic equations in the form (3.67) on the right-hand sides of the above equations, one easily obtains a second-order system for the Euler angles:

$$
\begin{pmatrix} \ddot{\Psi}_{\text{♁}} \\ \ddot{\varepsilon}_{\text{♁}} \\ \ddot{\Theta}_{\text{♁}} \end{pmatrix} = \begin{pmatrix} \left(-\sin\Theta_{\text{♁}}\,\dot{\omega}_{\text{♁}\mathcal{F}_1} - \cos\Theta_{\text{♁}}\,\dot{\omega}_{\text{♁}\mathcal{F}_2} + \dot{\varepsilon}_{\text{♁}}\,\dot{\Theta}_{\text{♁}} - \dot{\varepsilon}_{\text{♁}}\,\dot{\Psi}_{\text{♁}}\right) / \sin\varepsilon_{\text{♁}} \\ -\cos\Theta_{\text{♁}}\,\dot{\omega}_{\text{♁}\mathcal{F}_1} + \sin\Theta_{\text{♁}}\,\dot{\omega}_{\text{♁}\mathcal{F}_2} - \dot{\Psi}_{\text{♁}}\,\dot{\Theta}_{\text{♁}}\sin\varepsilon_{\text{♁}} \\ -\ddot{\Psi}_{\text{♁}}\cos\varepsilon_{\text{♁}} + \dot{\varepsilon}_{\text{♁}}\,\dot{\Psi}_{\text{♁}}\sin\varepsilon_{\text{♁}} + \dot{\omega}_{\text{♁}\mathcal{F}_3} \end{pmatrix} .
\tag{3.127}
$$

Equations (3.127) actually are second-order differential equations in the Euler angles, because the first derivatives of the components of the angular velocity vector $\boldsymbol{\omega}_{\text{♁}}$ in the Earth-fixed PAI-system only contain the components of this vector (which may in turn be written as functions of the Euler angles thanks to the kinematic equations (3.67)) and the components of the perturbing bodies in the Earth-fixed PAI-system, which are obtained by the corresponding components in the Earth-fixed PAI-system and the Euler angles as transformation parameters.

By replacing the subscript "♁" by the subscript "☾" in the above equations, one obtains the corresponding relations for the Moon.

When using the equations (3.127) instead of the first-order version (3.88), (3.68) one has to use the kinematic equations in the form (3.67) whenever the angular velocity vector is required.

For analytical investigations the first-order version of the equations is usually given the preference, when numerically solving the equations, the version (3.127) is better suited.

3.3.7 Kinematics of the Non-Rigid Earth

Strictly speaking, expressions like *body-fixed coordinate system*, *rotation* and *angular velocity* of a celestial body become meaningless when departing from the rigid-body model and allowing for deformations.

It is intuitively clear what has to be understood by the *deformation* of a celestial body: As opposed to the rigid body approximation one allows for the distances between individual mass elements of the body to vary in time. These variations are due to forces acting between the mass elements. The relevant deformations in the case of the Earth are small: the tidal deformations (mainly) due to Moon and Sun are, e.g., only of the order of a few decimeters on the Earth's surface and on the open ocean – an effect with a relative amplitude of about 10^{-7} when measured in Earth radii.

The center of mass of a deformable body is defined in the same way as that of a rigid body: The definition (3.55) may be taken over without any changes – one only has to keep in mind that the physical shape of the volume V occupied by the body may change in time. The same is true for the definition of a celestial body's inertia tensor: The definition (3.74) is suitable for a deformable body as well.

It is therefore still possible to write the position vector $\boldsymbol{x}_p$ (see Figure 3.2) of a particular volume element as the sum of the position vector of the Earth's center of mass $\boldsymbol{x}_{♁}$ and the geocentric position vector $\boldsymbol{r}_p$ of the volume element:

$$\boldsymbol{x}_p = \boldsymbol{x}_{♁} + \boldsymbol{r}_p \ . \tag{3.128}$$

In view of the fact that the deformations of an Earth-like planet are in general small, it makes sense to define a *rigid, rotating coordinate system* w.r.t. which the actual deformations of the body remain small at all times. Let us mention that such a system may not exist for all celestial bodies. In the case of the Sun there are, e.g., latitude dependent angular velocities of solar rotation, which would invalidate this concept.

Having introduced the rigid, rotating coordinate system, it is natural to associate the angular velocity vector $\boldsymbol{\omega}_{♁}$ with the rotating coordinate system, implying in turn that the three Euler angles $\varPsi_{♁}$, $\varepsilon_{♁}$ and $\varTheta_{♁}$ describe the orientation (attitude) of this rotating coordinate system in inertial space.

Equations (3.56) describe the transformation of the position vector of an arbitrary volume element of the rigid Earth between the inertial and the Earth-fixed PAI-systems. The same equations describe the transition between the inertial and the rigid, rotating system in the case of deformable bodies.

For a rigid Earth the velocity of a volume element of the Earth may be written as the sum of the velocity of the Earth's center of mass and the geocentric velocity of the element, which in turn is defined by the vector product (3.58). If we allow for deformations we have to use the more general relation

$$\dot{\boldsymbol{r}}_p = \boldsymbol{\omega}_{♁} \times \boldsymbol{r}_p + \delta\dot{\boldsymbol{r}}_p \ , \tag{3.129}$$

where $\delta\dot{\boldsymbol{r}}_p$ describes the motion of the volume element relative to the rotating system.

With the above interpretation of the Euler angles Euler's kinematic equations, i.e., the relations between the components of the angular velocity vector $\boldsymbol{\omega}_{\text{♁}}$ and the angular velocities $\dot{\Psi}_{\text{♁}}$, $\dot{\varepsilon}_{\text{♁}}$, $\dot{\Theta}_{\text{♁}}$, also may be used for the case of the non-rigid Earth.

The definitions (3.74) and (3.70) for the inertia tensor and the angular momentum may be used for non-rigid bodies as well. One has to keep in mind, however, that there is in general no coordinate system w.r.t. which all elements of the inertia tensor are constant in time.

The angular momentum of a deformable Earth is given by the definition (3.70), where the representation (3.129) for the velocities of the Earth's volume elements has to be used:

$$
\begin{aligned}
\boldsymbol{h}_{\text{♁}} &= \int\limits_{V_{\text{♁}}} \rho(\boldsymbol{r}_p)\ \boldsymbol{r}_p \times \dot{\boldsymbol{r}}_p\ dV_{\text{♁}}\ = \int\limits_{V_{\text{♁}}} \rho(\boldsymbol{r}_p)\ \boldsymbol{r}_p \times \{\, \boldsymbol{\omega}_{\text{♁}} \times \boldsymbol{r}_p + \delta\dot{\boldsymbol{r}}_p \,\}\ dV_{\text{♁}} \\
&= \mathbf{I}_{\mathcal{F}_e}\, \boldsymbol{\omega}_{\text{♁}}\ + \int\limits_{V_{\text{♁}}} \rho(\boldsymbol{r}_p)\ \boldsymbol{r}_p \times \delta\dot{\boldsymbol{r}}_p\ dV_{\text{♁}}\ =\ \mathbf{I}_{\text{♁}\mathcal{F}}\, \boldsymbol{\omega}_{\text{♁}}\ +\ \boldsymbol{\kappa}_{\text{♁}}\ ,
\end{aligned}
\tag{3.130}
$$

where $\mathbf{I}_{\text{♁}\mathcal{F}}$ is the Earth's inertia tensor expressed in a rigid, rotating coordinate system and $\boldsymbol{\kappa}_{\text{♁}}(t)$ is the angular momentum of the deformable Earth relative to the same system. Note that eqns. (3.130) are the generalized relations (3.73) between the angular momentum $\boldsymbol{h}$ of a deformable planet and the angular velocity vector $\boldsymbol{\omega}$ characterizing its rotation.

Up till now the definition of the rigid, rotating coordinate system was somewhat arbitrary: We just asked for a system relative to which the deformations would be small – if possible at all times. A particularly suitable rigid, rotating system may be defined by the requirement

$$
\boldsymbol{\kappa}_{\text{♁}}(t) \overset{\text{def}}{=} \mathbf{0}
\tag{3.131}
$$

for all times t. The idea of defining the rotating frame by asking the angular momentum due to deformations to vanish for all times is attributed to Félix Tissérand (1845–1896) (see [121]). The coordinate System $\mathcal{F}$ thus has a slightly different meaning in the case of the deformable Earth. It may no longer be defined as an Earth-fxed system, but as the system rotating with the rigid coordinate system, w.r.t. which there is no inner angular momentum.

Equations (3.131) are the condition equations for the realization of a Tissérand system. The actual realization of a Tissérand system is far from trivial, because we have no direct access to the velocities and densities in the Earth's interior. Usually, Tissérand systems are realized in a purely kinematic way using the coordinates and velocities of the space geodetic observing sites on the Earth's crust.

Tissérand systems are particularly suitable, because the equations of motion are formally very similar to those of the rigid body when expressed in this system.

3.3.8 Liouville-Euler Equations of Earth Rotation

For the derivation of the (orbital and rotational) equations of motion in the case of deformable bodies we have to depart from the basic equations of motion (3.81) for individual volume elements of Earth (and Moon) – as in the rigid-body case. These equations then have to be combined according to exactly the same pattern as in the case of the rigid body in order to obtain the equations of motion for the center of mass and for the angular momentum of the Earth.

When reviewing the derivation of the results (3.85) and (3.88) we observe, that we actually did not have to assume the rigidity of celestial bodies. We were only relying on the basic law *actio=reactio* between any two individual mass elements. As long as this assumption holds, the equations of motion (3.85), (3.88) and (3.90) remain the same for deformable bodies as for rigid bodies.

Naturally, one has to keep in mind that the angular momentum $\boldsymbol{h}_{♁}$ of the planet is now defined by eqn. (3.130), giving rise to the following equations for Earth rotation:

$$\frac{d}{dt}\left\{\mathbf{I}_{♁\mathcal{F}}\,\boldsymbol{\omega}_{♁}+\boldsymbol{\kappa}_{♁}\right\}=\boldsymbol{\ell}_{☾♁}+\boldsymbol{\ell}_{\odot♁}\ . \tag{3.132}$$

Equations (3.132) are vector equations. They may, however, also be interpreted as equations in the coordinates of the inertial system. Note that eqn. (3.132), defining the rotation of a non-rigid body in inertial space, and the first of eqns. (3.90), defining the same motion for a rigid body, are formally identical, if a Tissérand system is used in the former case. Observe also, however, that $\mathbf{I}_{♁\mathcal{F}}$ cannot be transformed into a system, where all of its elements are time independent.

The torques $\boldsymbol{\ell}_{\odot♁}$ and $\boldsymbol{\ell}_{☾♁}$ are defined by the integrals on the right-hand side of eqns. (3.88). When evaluating these integrals, one of course would have to take the deformations into account, as well. In view of the fact that the torques are small quantities, the rigid-body approximation is in practice good enough for the computation of these integrals.

In section 3.3 it was argued that the equations for the rotation of the Earth are particularly simple when referred to the body-fixed PAI-system. We cannot expect a comparable gain, when transforming the equations for Earth rotation into the rigid, rotating coordinate system – just because the inertia tensor will neither become diagonal nor time-invariant in this system.

For Earth-like planets it is, however, always possible to introduce a rigid, rotating coordinate system w.r.t. which the off-diagonal elements of the inertia tensor and the time-varying part of the diagonal elements are small when compared to the diagonal elements of the corresponding rigid body.

The transformation of eqns. (3.132) from the inertial into the rigid, rotating coordinate system follows the pattern of the transformation from the inertial into the Earth's PAI-system in the case of the rigid body (see section 3.3.4). After a few rather tedious, but elementary algebraic transformations one obtains:

$$\frac{d}{dt}\left\{\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{\omega}_{\earth}+\boldsymbol{\kappa}_{\earth}\right\}\;+\;\boldsymbol{\omega}_{\earth}\times\left\{\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{\omega}_{\earth}+\boldsymbol{\kappa}_{\earth}\right\}=\boldsymbol{\ell}_{\leftmoon\earth}+\boldsymbol{\ell}_{\odot\earth}\ , \tag{3.133}$$

where:

$$\begin{aligned}\boldsymbol{\ell}_{\leftmoon\earth}&=\frac{G\,m_{\leftmoon}}{r_{\leftmoon}^{5}}\,\boldsymbol{r}_{\leftmoon}\times\left(\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{r}_{\leftmoon}\right)\\ \boldsymbol{\ell}_{\odot\earth}&=\frac{G\,m_{\odot}}{r_{\odot}^{5}}\,\boldsymbol{r}_{\odot}\times\left(\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{r}_{\odot}\right)\ .\end{aligned} \tag{3.134}$$

When using a Tissérand system, equations (3.133) assume the following particularly simple form

$$\frac{d}{dt}\left\{\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{\omega}_{\earth}\right\}\;+\;\boldsymbol{\omega}_{\earth}\times\left\{\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{\omega}_{\earth}\right\}=\boldsymbol{\ell}_{\leftmoon\earth}+\boldsymbol{\ell}_{\odot\earth}\ . \tag{3.135}$$

Equations (3.135) are usually referred to as the *Liouville-Euler equations* of Earth rotation (named after Joseph Liouville (1809–1882) and Euler).

Note that for the special case of the rigid body eqns. (3.135) reduce to the simpler equations (3.124), *if* the PAI-system is used. For a rigid body we might use the Liouville-Euler equations (3.135) as the equations of motion referring to an arbitrary body-fixed system. In this case we might even make use of the fact that the inertia tensor is not a function of time and write

$$\mathbf{I}_{\earth\mathcal{F}}\,\dot{\boldsymbol{\omega}}_{\earth}\;+\;\boldsymbol{\omega}_{\earth}\times\left\{\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{\omega}_{\earth}\right\}=\boldsymbol{\ell}_{\leftmoon\earth}+\boldsymbol{\ell}_{\odot\earth}\ , \tag{3.136}$$

whereas we have to take into account the time variability of the inertia tensor for non-rigid Earth models:

$$\dot{\mathbf{I}}_{\mathcal{F}e}\,\boldsymbol{\omega}_{\mathcal{F}e}\;+\;\mathbf{I}_{\earth\mathcal{F}}\,\dot{\boldsymbol{\omega}}_{\earth}\;+\;\boldsymbol{\omega}_{\earth}\times\left\{\mathbf{I}_{\earth\mathcal{F}}\,\boldsymbol{\omega}_{\earth}\right\}=\boldsymbol{\ell}_{\leftmoon\earth}+\boldsymbol{\ell}_{\odot\earth}\ . \tag{3.137}$$

We are now in a position to solve the Liouville-Euler equations, *provided* the inertia tensor and its time derivative are known. For this purpose we have to specify the nature of the deformations considered. This discussion will be a central issue in Chapter II- 2.

3.4 Equations of Motion for an Artificial Earth Satellite

3.4.1 Introduction

The equations of motion for artificial Earth satellites and their derivation from the Newtonian axioms is closely related to the developments related to the Earth-Moon-Sun system. Figure 3.3 may also serve to describe the motion of an artificial Earth-satellite in its orbit around our planet. It reminds us, that in general, artificial Earth satellites should be viewed as "extended" objects. A complete description of the satellite's motion comprises the motion of its center of mass and the orientation of a satellite-fixed coordinate system (with origin in the center of mass of the satellite) in inertial space. In the context of artificial satellites this orientation w.r.t. inertial space is called the *attitude* of the satellite. Figure 3.5 illustrates the body-fixed coordinate system for a satellite of the US Global Positioning System.

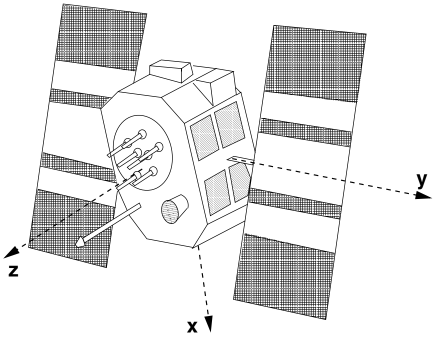


Fig. 3.5. Body-fixed coordinate system of a GPS satellite

Whereas the orientation of Earth and Moon is defined uniquely by Euler's equations (3.124), the same cannot be true for active satellites, like e.g., GPS satellites. These navigation satellites have to orient their antennas (along the positive z-axis) always (more or less) towards the Earth's center, in order to optimize signal reception for navigation and positioning on the Earth's surface or in the Earth-near space (precise orbit determination for LEOs). The solar panels of the GPS satellite provide the energy for the operation of the satellites. In order to optimize the energy gain, the panels' surfaces have to be perpendicular to the direction Sun-satellite at all times. This is why the *attitude* of the satellite has to be actively controlled. For GPS satellites the nominal attitude is maintained with *momentum wheels*, mounted on the axes of the satellites. Only occasionally, when the momentum wheels in the satellites are spinning too rapidly, the rotation of the wheels has to be stopped, and the attitude has to be maintained using thrusters (this de-spinning of the wheels is called "momentum dump"). If the attitude of satellites is maintained actively, the Euler equations for the rotation of a body are completely ruled out (or they should be modified to include the torques applied by these mechanisms).

There are satellites, for which the orientation is not important. The satellites Lageos I and Lageos II (Lageos II is shown in Figure 2.4 in Chapter 2) were, e.g., designed as massive, spherically symmetric bodies with Laser retro-reflectors distributed over the satellites' surface. As the name implies, the satellites were constructed to be observed by the Laser observation technique (for a precise definition of the Laser measurement see Chapter II- 3). The reduction of the distance measurement from the actual reflector to the satellite's center of mass is trivial and needs no knowledge of the attitude of the (spherically symmetric) satellite. Also, the mass distribution within the satellite is almost perfectly spherically symmetric, which is why the attitude of the satellite should not have any sizeable influence on the motion of the center of mass.

The two examples indicate that in satellite geodesy often the attitude plays a lesser role than the orbital motion, or, if the attitude is important, it is established by active control mechanisms onboard the satellite, momentum wheels and thrusters being the important tools for attitude maintenance. However, as soon as a satellite is not (or is no longer) actively controlled, its attitude may be derived from Euler's equations describing the rotation of a rigid body. This is why in section 3.4.3 we include the equations governing a rigid satellite's attitude under the assumptions that there are no active control mechanisms onboard. In the same section we also discuss the validity of separation of orbital and rotational motion for artificial satellites. In section 3.4.2 we uniquely focus uniquely on the motion of a satellite's center of mass.

3.4.2 Equations for the Center of Mass of a Satellite

Apart from the attitude-related issues mentioned in the introductory section 3.4.1 the following differences w.r.t. the equations of motion dealt with so far are relevant:

- The mass of an artificial satellite always may be neglected w.r.t to the masses of Earth, Moon, Sun, and planets. This aspect considerably reduces the complexity of the problem because we do not have to worry about the accelerations exerted by the satellite on these celestial bodies.
- The orbits and orientation of Earth, Sun and Moon, and planets, required to compute the satellite's motion, may be assumed as known.
- Due to the proximity of artificial Earth satellites with height above surface ranging from 150 – 200 km (below this height above the Earth's surface

an orbiting object will decay rather rapidly) up to, let us say, the geostationary belt at a geostationary distance of about 42000 km, it is no longer sufficient to take only the main term and the second-order terms of the Earth's gravitational potential into account. A much more complete description including hundreds of terms of the stationary part of the Earth's gravitational potential is needed.

- The tidal deformations of the Earth (solid Earth and ocean tides) have to be modeled as well when computing the gravitational attraction acting on the satellite.
- Non-gravitational forces like
  - Air drag due to the Earth's upper atmosphere,
  - solar radiation pressure effects,
  - thruster firings,
  - etc.

  have to be considered as well.

The first of the above aspects is the only one reducing the complexity of the problem. It allows us to address the equations for one satellite at the time, which is why usually only three differential equations (of second order) for the center of mass (and possibly three more equations describing the attitude) are considered subsequently. The second aspect does not pose delicate problems, in particular if the point mass approximation is used. The third aspect, the generalization of the Earth's gravitational potential, is the main topic of this section. Non-gravitational forces (aspect 5) will be dealt with in Chapter II- 3 together with the forces associated with the aspects of a non-rigid Earth. The discussion of tidal deformations is postponed to Chapter II- 3.

The Equations of Motion in the Inertial System. In the inertial system the equations of motion for an artificial satellite follow directly from the Newtonian axioms. In analogy to eqns. (3.81) the equations of motion of the center of mass $\boldsymbol{x}$ of a satellite of mass m may be set up as follows:

$$\begin{aligned} m\,\ddot{\boldsymbol{x}} = &- G\,m \int\limits_{V_{\text{♁}}} \rho_{p_r}\,\frac{\boldsymbol{x}-\boldsymbol{x}_p}{|\boldsymbol{x}-\boldsymbol{x}_p|^3}\,dV_{\text{♁}} \\ &- G\,m \sum_{j=1}^{n} m_j\,\frac{\boldsymbol{x}-\boldsymbol{x}_j}{|\boldsymbol{x}-\boldsymbol{x}_j|^3} \;+\; \sum \boldsymbol{f}_{\mathrm{ng}} \;+ \dots\;, \end{aligned} \tag{3.138}$$

where $\sum \boldsymbol{f}_{\mathrm{ng}}$ is the sum of non-gravitational forces acting on the satellite. Apart from the Earth's gravitational attraction the gravitational effects of $n \geq 2$ celestial bodies have to be modelled (Sun and Moon and possibly (other) planets), where the point mass model can be adopted for all perturbing bodies. This is justified in most cases due to the distances of the satellite w.r.t. these bodies. For very ambitious applications (very long satellite arcs) the term associated with lunar gravitation would have to be taken into account by a volume integral, as well.

Dividing all terms of the above equation by the mass of the satellite gives the acceleration of the satellite in the inertial reference frame. Using moreover the relative density function, we obtain the equation of motion of the satellite in the inertial system, with M = mass of the Earth:

$$\ddot{\boldsymbol{x}} = -GM \int\limits_{V_{\text{♁}}} \rho_{p_r} \, \frac{\boldsymbol{x} - \boldsymbol{x}_p}{|\boldsymbol{x} - \boldsymbol{x}_p|^3} \, dV_{\text{♁}} \; - \; G \sum_{j=1}^{n} m_j \, \frac{\boldsymbol{x} - \boldsymbol{x}_j}{|\boldsymbol{x} - \boldsymbol{x}_j|^3} \; + \; \sum \boldsymbol{a}_{\text{ng}} \; + \; \dots \,, \tag{3.139}$$

where $\boldsymbol{a}_{\text{ng}}$ are the non-gravitational accelerations (better: forces per mass unit) acting on the satellite. In the same approximation (i.e., point mass approximation to describe the relative motion of all celestial bodies except that of the satellite) the equations of motion for the Earth's center of mass may be written as (compare eqns. (3.82)):

$$\ddot{\boldsymbol{x}}_{\text{♁}} = -G \sum_{j=1}^{n} m_j \, \frac{\boldsymbol{x}_{\text{♁}} - \boldsymbol{x}_j}{|\boldsymbol{x}_{\text{♁}} - \boldsymbol{x}_j|^3} \; . \tag{3.140}$$

For highest accuracies one might use the equation (3.82) to model the motion of the Earth's center of mass. For all applications we consider in this book, the approximation (3.140) is sufficient. The generalization is straight forward and may be left to the reader.

The Equations of Motion in the Geocentric System. Subtracting eqn. (3.140) from eqn. (3.139) leads to the equation for the geocentric motion of the satellite' center of mass, $\boldsymbol{r} \overset{\text{def}}{=} \boldsymbol{x} - \boldsymbol{x}_{\text{♁}}$ (the geocentric position vectors for the other bodies are defined in the same way).

$$\begin{aligned} \ddot{\boldsymbol{r}} = & -GM \int\limits_{V_{\text{♁}}} \rho_{p_r} \, \frac{\boldsymbol{r} - \boldsymbol{r}_p}{|\boldsymbol{r} - \boldsymbol{r}_p|^3} \, dV_{\text{♁}} \\ & - G \sum_{j=1}^{n} m_j \left\{ \frac{\boldsymbol{r} - \boldsymbol{r}_j}{|\boldsymbol{r} - \boldsymbol{r}_j|^3} + \frac{\boldsymbol{r}_j}{r_j^3} \right\} \; + \; \sum \boldsymbol{a}_{\text{ng}} \; + \; \dots \; . \end{aligned} \tag{3.141}$$

The above equation is a vector equation. It may, however, also interpreted as an equation in Cartesian coordinates. Due to the fact that the equatorial plane is in an excellent approximation a plane of symmetry of the Earth, it makes sense to refer the above equations to the equatorial geocentric coordinate system of a particular reference epoch (which we always select as the system $J2000.0$). The system is a so-called quasi-inertial system (as it is not rotating w.r.t. any inertial system, but its origin is attached to the Earth's center of mass).

The gravitational terms on the right-hand sides of these equations may be written as gradients of a potential function.

$$\begin{aligned}\ddot{\boldsymbol{r}} = GM\,\nabla \int\limits_{V_{\text{♁}}} \frac{\rho_{p_r}}{|\boldsymbol{r}-\boldsymbol{r}_p|}\, dV_{\text{♁}} \\ + G\,\nabla \left[\sum_{j=1}^{n} m_j \left\{\frac{1}{|\boldsymbol{r}-\boldsymbol{r}_j|} + \frac{\boldsymbol{r}_j \cdot \boldsymbol{r}}{r_j^3}\right\}\right] + \sum \boldsymbol{a}_{\text{ng}} + \ldots \;.\end{aligned} \tag{3.142}$$

Note that tidal effects, which are due to the Earth's deformation caused by the gravitational attraction of Moon and Sun (see Chapter II- 2), are included in the equations of motion (3.142) provided the volume occupied by the Earth's body and the density ρ_{p_r} are considered as time-varying quantities. In order to simplify the discussion, we will assume for the remainder of this chapter that the Earth is rigid and postpone the discussion of the tidal potential to Chapters II- 2 and II- 3.

It was already mentioned that the equations of motion (3.142) may also be interpreted as equations in the components of vector $\boldsymbol{r}$ in the geocentric quasi-inertial system. It is, however, much more convenient to evaluate the volume integral in the Earth's PAI-system and to evaluate the gradient in this system. This is achieved by transforming the component matrix from the inertial to the Earth's PAI-system, taking the gradient in this system system, and then by transforming the gravitational acceleration due to the Earth into the inertial system. Using this concept, the equations of motion in the geocentric quasi-inertial system may be written in the form

$$\ddot{\boldsymbol{r}} = GM\,\mathbf{T}_{\text{♁}}\;\nabla V(\boldsymbol{r}) \;+\; G\nabla \left[\sum_{j=1}^{n} m_j \left\{\frac{1}{|\boldsymbol{r}-\boldsymbol{r}_j|} + \frac{\boldsymbol{r}_j \cdot \boldsymbol{r}}{r_j^3}\right\}\right] + \sum \boldsymbol{a}_{\text{ng}} + \ldots\,, \tag{3.143}$$

with the understanding that all component matrices in the equations of motion (3.143) refer to the inertial system, except for the term $\nabla V(\boldsymbol{r})$, which refers to the Earth's PAI-system. The back-transformation into the inertial system is performed by the matrix $\mathbf{T}_{\text{♁}}$, eqn. (3.56), which is why $\mathbf{T}_{\text{♁}}\;\nabla V(\boldsymbol{r})$ is a component matrix referring to the quasi-inertial geocentric system (as all other terms on the right-hand side of eqn. (3.143)).

The Earth's Stationary Gravitational Potential. According to eqn. (3.143) the potential function of the Earth may be written as

$$V(\boldsymbol{r}) = GM \int\limits_{V_{\oplus}} \frac{\rho_{p_r}}{|\boldsymbol{r}-\boldsymbol{r}_p|}\, dV_{\oplus} \; . \tag{3.144}$$

In this definition $V(\boldsymbol{r})$ is always positive. One easily verifies that the following equation holds by taking first the gradient $\nabla V(\boldsymbol{r})$ of the above scalar function, then the divergence $\nabla \cdot \nabla V(\boldsymbol{r}) \overset{\text{def}}{=} \Delta V(\boldsymbol{r})$ of the gradient:

$$\Delta V(\boldsymbol{r}) \overset{\text{def}}{=} \left\{ \frac{\partial^2}{\partial r_1^2} + \frac{\partial^2}{\partial r_2^2} + \frac{\partial^2}{\partial r_3^2} \right\} V(\boldsymbol{r}) = 0 \; , \tag{3.145}$$

where $r_i\,,\ i = 1,2,3$, are the Cartesian coordinates of the vector $\boldsymbol{r}$ in the Earth-fixed system (the equation would of course hold in any Cartesian coordinate system). Equation (3.145) is called the *Laplace equation.* This equation holds for any potential for vectors $\boldsymbol{r}$ outside the mass distribution (for satellites, as long as they are in orbit(!), this condition is certainly met).

As the mass distribution within the Earth is close to spherically symmetric, it makes sense to express Laplace's equation in spherical, rather than rectangular coordinates. Let us introduce the following spherical coordinates in the Earth-fixed system:

$$\begin{aligned}
r_1 &= r\cos\phi\cos\lambda \; ; & r &= \sqrt{r_1^2+r_2^2+r_3^2} \\
r_2 &= r\cos\phi\sin\lambda \; ; & \phi &= \arcsin\frac{r_3}{\sqrt{r_1^2+r_2^2+r_3^2}} \\
r_3 &= r\sin\phi \; ; & \lambda &= \arctan\frac{r_2}{r_1} \; ,
\end{aligned} \tag{3.146}$$

where r is the geocentric distance, ϕ the latitude, and λ the longitude of the satellite. ϕ and λ are geocentric coordinates, where the longitude λ is positive East of the Greenwich meridian. For points on the northern hemisphere the latitude ϕ is positive, for the southern hemisphere it is negative.

It is a tedious, but straightforward exercise to transform the Laplace equation from rectangular to spherical coordinates:

$$\left\{ \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial}{\partial\phi}\right) + \frac{1}{r^2\cos^2\phi}\frac{\partial^2}{\partial\lambda^2} \right\} V(\boldsymbol{r}) = 0 \; , \tag{3.147}$$

where

$$\Delta = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial}{\partial\phi}\right) + \frac{1}{r^2\cos^2\phi}\frac{\partial^2}{\partial\lambda^2} \tag{3.148}$$

is the *Laplace operator* in spherical coordinates.

Every solution of the Laplace equation, be it expressed in rectangular (3.145) or in spherical coordinates (3.147), is called a *harmonic function.* Solutions of eqn. (3.147) are called spherical harmonic functions, or briefly *spherical harmonics.*

It is particularly attractive that the solutions of the second order partial differential equation (3.147) may be separated in spherical coordinates:

$$V(\boldsymbol{r}) = R(r)\,\varPhi(\phi)\,\varLambda(\lambda) \ . \tag{3.149}$$

A formal proof for this equation and the derivation of its solution may be found in any treatment of potential theory and in many textbooks of physical geodesy. Let us mention in particular the very concise treatment by Kaula [62] and the more elaborate and complete treatments in [52] and [124]. The final result may be brought into the following form:

$$V(r,\lambda,\phi) = \frac{GM}{r}\sum_{i=0}^{\infty}\left(\frac{a_{\text{♁}}}{r}\right)^{i}\sum_{k=0}^{i}P_i^k(\sin\phi)\big\{\,C_{ik}\cos k\lambda + S_{ik}\sin k\lambda\,\big\}\ , \tag{3.150}$$

where $a_{\text{♁}} \approx 6378137$ m (see Table II- 2.1) is the equatorial radius of the Earth. The functions $P_i^k(x)$ are the *associated Legendre functions*, which may be defined as follows:

$$\begin{aligned} P_i^0(x) &= P_i(x) = \frac{1}{2^i\, i!}\,\frac{d^i}{dx^i}\Big\{(x^2-1)^i\Big\} \\ P_i^k(x) &= (1-x^2)^{\frac{k}{2}}\,\frac{d^k}{dx^k}\big\{P_i(x)\big\}\ , \quad k = 0,1,\ldots,i\ , \end{aligned} \tag{3.151}$$

where $P_i(x)$ are the *Legendre polynomials* as they were already introduced in eqns. (3.101). The coefficients C_{ik} and S_{ik} are defined as (see, e.g., [124]):

$$\begin{aligned} C_{i0} &= \frac{1}{a_{\text{♁}}^i}\int\limits_{V_{\text{♁}}}\rho_{p_r}\, r^i\, P_i(\sin\phi_p)\ dV\ ; \quad i \ge 0 \\ C_{ik} &= \frac{2}{a_{\text{♁}}^i}\,\frac{(i-k)!}{(i+k)!}\int\limits_{V_{\text{♁}}}\rho_{p_r}\, r^i\, P_i^k(\sin\phi_p)\cos k\lambda_p\ dV\ ; \quad i,k \ge 0\ ,\ k \le i \\ S_{ik} &= \frac{2}{a_{\text{♁}}^i}\,\frac{(i-k)!}{(i+k)!}\int\limits_{V_{\text{♁}}}\rho_{p_r}\, r^i\, P_i^k(\sin\phi_p)\sin k\lambda_p\ dV\ ; \quad i,k > 0\ ,\ k \le i\ . \end{aligned} \tag{3.152}$$

(Remember that $\rho_{p_r} = \rho(\boldsymbol{r}_p)/M$ is the relative density as introduced in eqns. (3.83)). The use of the equatorial radius $a_{\text{♁}}$ of the Earth and the isolation of the factor $\frac{GM}{r}$ in the representation (3.150) for the Earth's gravitational potential function is somewhat arbitrary. It has the advantage that the coefficients C_{ik} and S_{ik} become dimensionless. Another advantage of this scaling procedure resides in the fact that for LEOs, where $\frac{a_{♁}}{r} \approx 1$, the coefficients may be easily compared with each other, in particular with the main term $C_{00} = 1$.

The index i is called the *degree*, k the *order* of the spherical harmonic function in the above notation. For the sake of completeness we mention that the Legendre functions, multiplied by either $\sin k\lambda$ or $\cos k\lambda$, are called *spherical functions*; if multiplied in addition with $\frac{1}{r^{i+1}}$ they are called *spherical harmonic functions* (and they are solutions of the Laplace equation (3.147)).

It is worthwhile to calculate the first few low degree (and order) coefficients C_{ik} and S_{ik} of the development (3.150). It should be mentioned that, because of

$$\sin k\lambda = 0 \ , \qquad \text{for } k = 0 \ ,$$

one may set

$$S_{i0} = 0 \ , \qquad \text{for } i = 0, 1, 2, \dots \ . \tag{3.153}$$

Let us now calculate all terms up to degree $i = 1$ using the defining equations (3.152). In order to preserve generality the results will first be given w.r.t. to an arbitrary coordinate system, only in the last step it will be assumed that the coordinate system refers to the center of mass. The result for the orders zero and one are:

$$\begin{aligned}
C_{00} &= \int\limits_{V_{♁}} \rho_{p_r} \, r_p^0 \; dV = 1 \\
C_{10} &= \frac{1}{a_{♁}} \int\limits_{V_{♁}} \rho_{p_r} \, r_p \sin\phi_p \; dV = \frac{1}{a_{♁}} \int\limits_{V_{♁}} \rho_{p_r} \, r_{p_3} \; dV = \frac{r_3}{a_{♁}} = 0 \\
C_{11} &= \frac{1}{a_{♁}} \int\limits_{V_{♁}} \rho_{p_r} \, r_p \cos\phi_p \cos\lambda_p \; dV = \frac{1}{a_{♁}} \int\limits_{V_{♁}} \rho_{p_r} \, r_{p_1} \; dV = \frac{r_1}{a_{♁}} = 0 \\
S_{11} &= \frac{1}{a_{♁}} \int\limits_{V_{♁}} \rho_{p_r} \, r \cos\phi_p \sin\lambda_p \; dV = \frac{1}{a_{♁}} \int\limits_{V_{♁}} \rho_{p_r} \, r_{p_2} \; dV = \frac{r_2}{a_{♁}} = 0 \ .
\end{aligned} \tag{3.154}$$

The equalities "= 0" hold, because the coordinate system refers to the Earth's center of mass. The term C_{00}, multiplied by M, is the total mass of the Earth, the coefficients C_{10}, C_{11} and S_{11} are the center of mass coordinates divided by $a_{♁}$ (expressed in units of the equatorial radius $a_{♁}$).

In principle the origin of the coordinate system should coincide with the center of mass of the Earth (this is what we assumed so far). In practice, one has the difficulty that the center of mass of the Earth has to be determined from satellite geodetic observations relative to a polyhedron of observing sites on the Earth's crust. This implies that, with improving observation techniques and processing strategies, better and better estimates for the center of mass relative to the polyhedron of observing stations become available. If a new center of mass estimate becomes available, one has in principle two options:

(a) Coefficients $C_{10} \neq 0,\ C_{11} \neq 0$ und $S_{11} \neq 0$ are accepted, or,

(b) $C_{10} = C_{11} = S_{11} = 0$ is enforced, but then the coordinates of the entire polyhedron of observing sites have to be adapted in order to refer to the newly established center of mass.

In practice one often prefers option (a) – obviously there are more users of station coordinates than of potential coefficients.

The coefficients of second degree may all be expressed in terms of the elements of the inertia tensor. The developments are first given w.r.t. an arbitrary coordinate system; in the last step the result refers to the PAI-system.

$$\begin{aligned}
C_{20} &= \frac{1}{a_{\text{♁}}^2} \int\limits_{V_{\text{♁}}} \rho_{p_r}\, r_p^2 \left[\, \sin^2\phi_p \;-\; \tfrac{1}{2}\cos^2\phi_p \,\right] dV \\
&= \frac{1}{a_{\text{♁}}^2} \int\limits_{V_{\text{♁}}} \rho_{p_r} \left[\, r_{p_3}^2 \;-\; \tfrac{1}{2}\left(r_{p_1}^2 + r_{p_2}^2\right) \right] dV \\
&= \frac{1}{M\, a_{\text{♁}}^2} \left[\, \tfrac{1}{2}\left(I_{\text{♁}\mathcal{F}_{11}} + I_{\text{♁}\mathcal{F}_{22}}\right) - I_{\text{♁}\mathcal{F}_{33}} \,\right] = \frac{1}{M\, a_{\text{♁}}^2} \left[\, \tfrac{1}{2}\left(A_{\text{♁}} + B_{\text{♁}}\right) - C_{\text{♁}} \,\right] \\[2ex]
C_{21} &= \frac{1}{a_{\text{♁}}^2} \int\limits_{V_{\text{♁}}} \rho_{p_r}\, r_p^2 \,\cos\lambda_p\, \cos\phi_p\, \sin\phi_p\; dV \\
&= \frac{1}{a_{\text{♁}}^2} \int\limits_{V_{\text{♁}}} \rho_{p_r}\, r_{p_1}\, r_{p_3}\; dV = -\frac{1}{M\, a_{\text{♁}}^2}\; I_{\text{♁}\mathcal{F}_{13}} \;=\; 0 \\[2ex]
S_{21} &= \frac{1}{a_{\text{♁}}^2} \int\limits_{V_{\text{♁}}} \rho_{p_r}\, r_p^2 \,\sin\lambda_p\, \cos\phi_p\, \sin\phi_p\; dV \\
&= \frac{1}{a_{\text{♁}}^2} \int\limits_{V_{\text{♁}}} \rho_{p_r}\, r_{p_2}\, r_{p_3}\; dV = -\frac{1}{M\, a_{\text{♁}}^2}\; I_{\text{♁}\mathcal{F}_{23}} \;=\; 0
\end{aligned} \tag{3.155}$$

$$\begin{aligned}
C_{22} &= \frac{1}{a_{♁}^2} \int\limits_{V_{♁}} \rho_{p_r}\, \tfrac{1}{4}\, R^2\, \cos^2\phi_p \left(\cos^2\lambda_p - \sin^2\lambda_p\right)\, dV \\
&= \frac{1}{a_{♁}^2} \int\limits_{V_{♁}} \rho_{p_r}\, \tfrac{1}{4} \left(r_{p_1}^2 - r_{p_2}^2\right)\, dV \\
&= \frac{1}{4}\, \frac{1}{M\, a_{♁}^2} \left(I_{♁\mathcal{F}_{22}} - I_{♁\mathcal{F}_{11}}\right) = \frac{1}{4}\, \frac{1}{M\, a_{♁}^2}\, \left(B_{♁} - A_{♁}\right) \\
S_{22} &= \frac{1}{a_{♁}^2} \int\limits_{V_{♁}} \rho_{p_r}\, \tfrac{1}{2}\, R^2\, \sin\lambda_p\, \cos\lambda_p\, \cos^2\phi_p\, dV \\
&= \frac{1}{a_{♁}^2} \int\limits_{V_{♁}} \rho_{p_r}\, \tfrac{1}{2}\, r_{p_1}\, r_{p_2}\, dV = -\frac{1}{2}\, \frac{1}{M\, a_{♁}^2}\, I_{♁\mathcal{F}_{12}} = 0\ .
\end{aligned}$$

From the above results it may be seen that, only if the coordinate axes coincide with the axes of principal inertia, we have

$$C_{21} = S_{21} = S_{22} = 0\ . \tag{3.156}$$

It is an old tradition to let the first axis of the Earth-fixed coordinate system lie in the Greenwich meridian. This is why in practice we might have $C_{21} = S_{21} = 0$, but $S_{22} \neq 0$.

If only terms up to degree 2 are taken into account and if a geocentric coordinate system with its third axis lying in the figure axis of the Earth is chosen, the potential function (3.150) assumes the following simple form:

$$\begin{aligned}
V(r,\lambda,\phi) = \frac{GM}{r} \;+\; \frac{GM}{M\, r^3} \Big\{ &\left(\tfrac{3}{2}\sin^2\phi \;-\; \tfrac{1}{2}\right) \left[\, \tfrac{1}{2}\, \left(A_{♁} + B_{♁}\right) - C_{♁}\, \right] \\
&+ 3\cos^2\phi \left[\, \tfrac{1}{4}(B_{♁} - A_{♁})\cos 2\lambda \;-\; \tfrac{1}{2}\, I_{♁\mathcal{F}_{12}}\, \sin 2\lambda \right] \Big\}\ .
\end{aligned} \tag{3.157}$$

Except for the term proportional to $I_{♁\mathcal{F}_{12}}$, which would disappear if the coordinate system would coincide with the Earth's three principal axes of inertia, the formula is just another version of MacCullagh's formula (3.105) .

Let us now discuss the general development (3.150) of the Earth's potential function. Terms of any degree i and of order $k = 0$ do not depend on the longitude λ. Their latitude dependence is defined by the Legendre polynomials $P_i(\sin\phi)$. Legendre polynomials of degree i have exactly i different roots in the interval $I = [-1, +1]$. Relating this to the concrete problem we distinguish exactly $i + 1$ latitude zones on the unit sphere, at the borders of which the Legendre polynomial changes sign. Figure 3.6 illustrates the case $i = 6,\ k = 0$, where zones with positive polynomial values are white, those with negative values black. Terms with $k = 0$ of the potential function are called *zonal* terms. According to their definition (3.151), the associated Legendre functions $P_i^i(\sin\phi)$ are constants, i.e., they neither depend on latitude

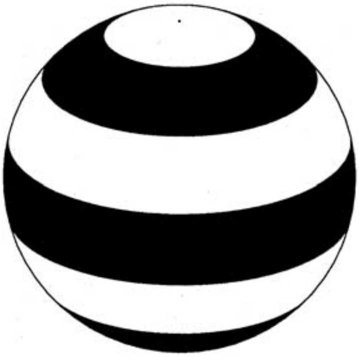


Fig. 3.6. Zonal harmonic function $(i = 6\,,\ k = 0)$

nor longitude. The associated spherical functions are either multiplied by $\sin i\lambda$ or $\cos i\lambda$. Therefore they have the same sign for all latitudes, but they change sign at i equally spaced longitudes. This is why terms with $i = k$ are called *sectorial harmonics.* Figure 3.7 illustrates a sectorial harmonic function with $i = k = 7$, where zones of identical sign of the harmonic function have the same shading. Harmonic functions with $k \neq 0$ and $k \neq i$ are called *tesseral functions.* Tesseral functions divide the sphere into $2k \cdot (i - k)$ different regions, in $k \cdot (i - k)$ of which the tesseral function assumes a positive and in the others negative values. Figure 3.8 illustrates a term of degree $i = 13$ and order $k = 7$.

In the development (3.150) one often uses the fully normalized Legendre functions $\bar{P}_i^k(\sin\phi)$. The normalization is done according to the following scheme:

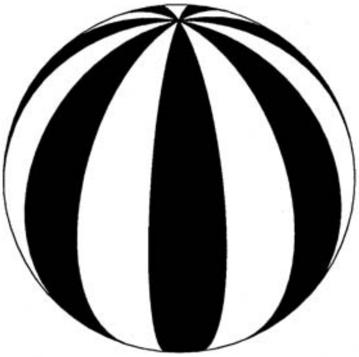


Fig. 3.7. Sectorial harmonic function $(i = 7\,,\ k = 7)$

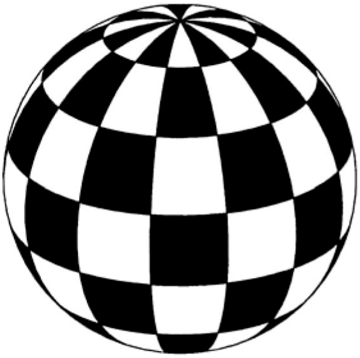


Fig. 3.8. Tesseral harmonic function $(i = 13\,,\ k = 7)$

$$\begin{aligned}
C_{ik} \quad &= \sqrt{\frac{2\,(2i+1)\,(i-k)!}{(i+k)!}}\ \bar{C}_{ik}\ , \quad k > 0 \\
C_{i0} \quad &= \sqrt{2i+1}\ \bar{C}_{i0} \\
S_{ik} \quad &= \sqrt{\frac{2\,(2i+1)\,(i-k)!}{(i+k)!}}\ \bar{S}_{ik}\ , \quad k > 0 \\
\bar{P}_i^k(\sin\phi) &= \sqrt{\frac{2\,(2i+1)\,(i-k)!}{(i+k)!}}\ P_i^k(\sin\phi)\ , \quad k > 0 \\
\bar{P}_i^0(\sin\phi) &= \sqrt{2i+1}\ P_i^0(\sin\phi)\ .
\end{aligned} \tag{3.158}$$

Table 3.1 contains the coefficient, complete up to degree and order 4, of the development (3.150) of the JGM3 (Joint Gravity Model 3) [120]. The coefficient $\bar{C}_{20}$, characterizing the flattening of the Earth, dominates clearly. There is no obvious hierarchy in the coefficients after $\bar{C}_{20}$. It makes therefore sense to speak of a *flattened* or *oblate* Earth; but to use terms like "pear-shaped Earth" is perhaps slightly exaggerated.

The adopted values for GM and $a_♁$ are scaling constants of the gravitational model, which is why the values for GM and $a_♁$ used in $JGM3$ are contained in Table 3.1. There are no terms of first degree, indicating that the origin of the terrestrial coordinate system underlying $JGM3$ is the Earth's center of mass. The coefficients $\bar{C}_{21}$ and $\bar{S}_{21}$ are very small, implying that the third axis very closely coincides with the figure axis of the Earth. The numerical value for $\bar{S}_{22}$ is comparatively big, due to the fact that no attempt was made to let the equatorial axes coincide with the axes of the second and third principal moments of inertia. The numerical value for $\bar{C}_{22}$ shows that the mass distribution in the Earth is not fully rotationally symmetric.

Table 3.1. The first terms of the Earth gravity model *JGM3*

Coefficient	Value	Coefficient	Value
GM	$398.60044150 \cdot 10^{12}\ \mathrm{m^3 s^{-2}}$	$a_{♁}$	6378136.30 m
$\bar{C}_{20}$	$-0.48416954845647 \cdot 10^{-3}$	$\bar{S}_{20}$	
$\bar{C}_{21}$	$-0.18698764000000 \cdot 10^{-9}$	$\bar{S}_{21}$	$+0.11952801000000 \cdot 10^{-8}$
$\bar{C}_{22}$	$0.24392607486563 \cdot 10^{-5}$	$\bar{S}_{22}$	$-0.14002663975880 \cdot 10^{-5}$
$\bar{C}_{30}$	$+0.95717059088800 \cdot 10^{-6}$	$\bar{S}_{30}$	
$\bar{C}_{31}$	$+0.20301372055530 \cdot 10^{-5}$	$\bar{S}_{31}$	$+0.24813079825561 \cdot 10^{-6}$
$\bar{C}_{32}$	$+0.90470634127291 \cdot 10^{-6}$	$\bar{S}_{32}$	$-0.61892284647849 \cdot 10^{-6}$
$\bar{C}_{33}$	$+0.72114493982309 \cdot 10^{-6}$	$\bar{S}_{33}$	$+0.14142039847354 \cdot 10^{-5}$
$\bar{C}_{40}$	$+0.53977706835730 \cdot 10^{-6}$	$\bar{S}_{40}$	
$\bar{C}_{41}$	$-0.53624355429851 \cdot 10^{-6}$	$\bar{S}_{41}$	$-0.47377237061597 \cdot 10^{-6}$
$\bar{C}_{42}$	$+0.35067015645938 \cdot 10^{-6}$	$\bar{S}_{42}$	$+0.66257134594268 \cdot 10^{-6}$
$\bar{C}_{43}$	$+0.99086890577441 \cdot 10^{-6}$	$\bar{S}_{43}$	$-0.20098735484731 \cdot 10^{-6}$
$\bar{C}_{44}$	$-0.18848136742527 \cdot 10^{-6}$	$\bar{S}_{44}$	$+0.30884803690355 \cdot 10^{-6}$

We stated above that the coefficients do not show a clear hierarchy. There is, however, an order of magnitude rule, called *Kaula's rule of thumb* [62], stating that

- the quantity
$$\sigma_i^2 \overset{\text{def}}{=} \sum_{k=0}^{i} \left[\bar{C}_{ik}^2 + \bar{S}_{ik}^2 \right] , \tag{3.159}$$
- which may be viewed as the power spectral density of the degree variances of degree i,
- corresponding to a half wavelength of
$$l_i \approx \frac{2\pi a_{♁}}{2i} \approx \frac{20'000}{i}\ \text{km} \tag{3.160}$$
on the Earth surface,
- decreases according to the rule
$$\sigma_i^2 = \frac{160 \cdot 10^{-12}}{i^3} . \tag{3.161}$$

Kaula [62] spelled out his rule at a time when the geopotential was not yet well established by satellite geodesy, which is why he mainly used terrestrial gravimetry data for his assessment. One may ask the question why σ_i^2 is considered as a spectral power density: Equation (3.150) represents the potential function on the Earth surface (more precisely for $r \overset{\text{def}}{=} a_{♁}$) as a linear combination of periodic functions. All terms of the same degree approximately have the same wavelength (as stated above). The power contained in a periodic signal is, on the other hand, proportional to the square of the amplitude of

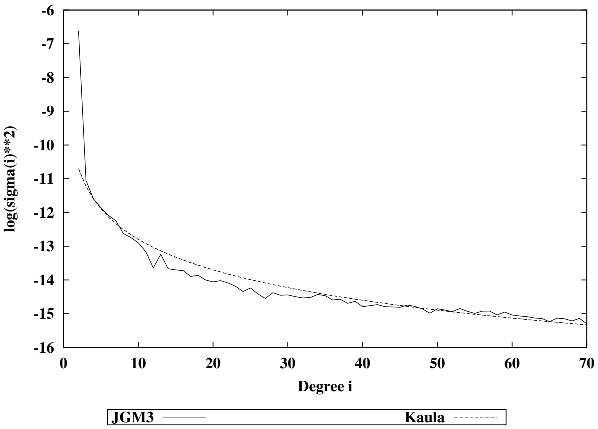


Fig. 3.9. Power spectral density of the Earth's gravitational field as a function of the spherical harmonics of degree i according to Kaula's rule and using the coefficients of $JGM3$

the signal. Therefore we may interpret the quantity σ_i^2 as a measure for the power contained in all terms of degree i of the gravitational field. Figure 3.9 illustrates that Kaula's rule of thumb is an excellent order of magnitude estimate. Figure 3.9 shows the logarithm of σ_i^2 for $i = 2, 3, \ldots, 70$, as a function of the degree i using the model $JGM3$. The Joint Gravity Model is one of the most recent global models for the Earth's gravitational field. It is based to a large extent on satellite geodetic observations. For an overview of the development of Earth gravitational models since the beginning of the Space Age we refer to [75].

Let us conclude this section by representing the Earth's potential function (3.150) in a slightly different way. Using the definitions

$$\begin{aligned} C_{ik} &\overset{\text{def}}{=} -J_{ik}\cos k\lambda_{ik} \\ S_{ik} &\overset{\text{def}}{=} -J_{ik}\sin k\lambda_{ik} \\ J_{ik} &= \sqrt{C_{ik}^2 + S_{ik}^2} \\ k\,\lambda_{ik} &= \arctan\left(\frac{-S_{ik}}{-C_{ik}}\right) \ , \end{aligned} \tag{3.162}$$

the potential function of the Earth may be written in a form which is (sometimes) better suited for studying the perturbations of an artificial Earth satellite:

$$V(r,\lambda,\beta) = \frac{GM}{r} - \frac{GM}{r}\sum_{i=2}^{\infty}\left(\frac{a_{\oplus}}{r}\right)^i \sum_{k=0}^{i} P_i^k(\sin\beta)\, J_{ik}\cos k(\lambda - \lambda_{ik}) \ . \tag{3.163}$$

Note that the sign associated with the perturbation potential is purely conventional and that the terms of degree 1 were assumed to be zero.

3.4.3 Attitude of a Satellite

Coupling of Orbital and Rotational Motion. So far it was assumed that orbital and rotational motion of a satellite may be established independently. We considerably extended the model for the motion of the satellite's center of mass by including not only Earth gravity terms up to the second order, as in eqns. (3.118) describing the geocentric orbital motion of Sun and Moon, but also up to any order that might seem appropriate; on the other hand, the term due to the finite size of the satellite was neglected.

Let us check whether this procedure was justified by transcribing the eqns. for the Moon in eqns. (3.118) to a satellite, and by taking into account only the (generalized) two-body orbital accelerations of the Earth-satellite system. Under these assumptions eqns. (3.118) read as

$$
\begin{aligned}
\ddot{\boldsymbol{r}} = & - GM \, \frac{\boldsymbol{r}}{r^3} \\
& + \frac{GM}{2\,M} \, \mathbf{T}_{\text{♁}} \, \nabla \left\{ \frac{A_{\text{♁}} + B_{\text{♁}} + C_{\text{♁}}}{r^3} - \frac{3 \, \left(r_1^2 A_{\text{♁}} + r_2^2 B_{\text{♁}} + r_3^2 C_{\text{♁}} \right)}{r^5} \right\} \\
& + \frac{GM}{2\,m} \, \mathbf{T}_s \, \nabla \left\{ \frac{A_s + B_s + C_s}{r^3} - \frac{3 \, \left(r_1^2 A_s + r_2^2 B_s + r_3^2 C_s \right)}{r^5} \right\} \, .
\end{aligned}
\tag{3.164}
$$

The satellite's mass m was neglected in eqns. (3.164) in the sum of masses $M + m$. In view of the Earth's mass of $M \approx 6 \cdot 10^{24}$ kg (Table II- 2.1) this is certainly justified. Equations (3.164) are correct up to terms of second order. The model of the acceleration due to the mass distribution of the Earth is contained in the second term (first bracket). This approximation was much improved previously by generalizing the term proportional to $\frac{GM}{2M}$ to any degree and not only the second as in the above equations.

The last term proportional to $\frac{GM}{2m}$ in eqn. (3.164) was not considered in the equations of motion previously established. Whether or not this was justified, depends to a large extent on the moments of inertia, more precisely on the quantities A_s/m, B_s/m, and C_s/m of the satellite. Obviously the term is nearly zero, for $A_s = B_s = C_s$. For (close to) spherically symmetric satellites our procedure was therefore correct.

Let us consider now a satellite consisting of two equal point masses of mass $m/2$ separated by a rigid, thin rod (dumb-bell shaped satellite). Clearly, knowledge of the satellite's attitude is required in the second bracket (last term). Let us assume that the rod has a length of $l = 20$ m. This interesting construction has the following principal moments of inertia

$$
A = 0 \; ; \quad B = C = 2 \, \frac{l^2}{4} \, m = 2 \cdot 100 \cdot m \; [\mathrm{m}^2\mathrm{kg}] \; . \tag{3.165}
$$

In the least favourable situation (i.e., when the neglected term assumes maximum value), the rod points to the center of the Earth and only the first term in the second pair of brackets in eqns. (3.164) contributes to the perturbing acceleration $\boldsymbol{a}$, where:

$$|\boldsymbol{a}| < 2\,\frac{3\,GM}{2\,m\,r^4} \cdot 2\,\frac{l^2}{4}\,m = \frac{3\,l^2\,GM}{2\,r^4} \approx 3.6 \cdot 10^{-11}\ [\mathrm{m/s^2}]\ . \tag{3.166}$$

This looks like a small acceleration, but, in view of the fact that the smallest perturbing accelerations considered today are of the order of 10^{-12} m s^{-2}, it would have to be taken into account. Because

- the Earth radius was used for the estimate (3.166) (for satellites in GPS orbits we would have $r \approx 26500$ km, which would reduce the estimate by a factor of about 300, rendering the term completely irrelevant), that
- a very special set of principal moments of inertia ($A_s = 0$, $B_s = C_s = 2\,\frac{l^2}{2^2}\,m$) was used for the satellite, and that
- the arc lengths spanned by one set of initial conditions usually are not longer than a few days,

we conclude that in practice the decoupling of orbital and rotational motion is in general justified for artificial Earth satellites.

The situation is different, if, e.g., the orbit of a tethered satellite is considered, where the connection between the two "point masses" may have a length l of several km. The orbital and rotational motion of such constructions is much more complicated than the cases considered here – in particular because the assumption of a rigid body does no longer hold and because surface forces (drag and radiation pressure) become rather important. Probably it is safe to state that the term discussed here (and neglected previously and subsequently) would be the least problematic.

Attitude of a Satellite. The above considerations show that in a good approximation the attitude of a passive satellite may be determined by Euler's equations under the assumption that the orbital motion of the satellite's center of mass is known. The derivation of the equations describing the attitude of a rigid satellite, which shall be characterized by the three principal moments of inertia A_s, B_s, and C_s, is done in close analogy to the derivation of the corresponding equations for Earth and Moon rotation. The equations for the rotation of the satellite in a satellite-fixed coordinate system with the origin in the center of mass and the coordinate axes as axes of principal inertia may be transcribed from the corresponding equations (3.124) and (3.68) of Earth rotation:

$$\begin{pmatrix} \dot{\omega}_{s1} \\ \dot{\omega}_{s2} \\ \dot{\omega}_{s3} \end{pmatrix} + \begin{pmatrix} \gamma_{s1}\omega_{s2}\omega_{s3} \\ \gamma_{s2}\omega_{s3}\omega_{s1} \\ \gamma_{s3}\omega_{s1}\omega_{s2} \end{pmatrix} = +\,\frac{3\,GM}{r^5} \begin{pmatrix} \gamma_{s1} r_2 r_3 \\ \gamma_{s2} r_3 r_1 \\ \gamma_{s3} r_1 r_2 \end{pmatrix}\ . \tag{3.167}$$

In the transition from eqns. (3.124) to eqns. (3.167) only the torque due to the Earth was taken into account. The torques due to the Moon and the Sun are orders of magnitude smaller and are not considered here. Note that the quantities γ_{s_i}, $i = 1, 2, 3$, are defined in analogy to the corresponding quantities (3.123) related to Earth and Moon. $\boldsymbol{\omega}_s$ is the angular velocity vector of the satellite, its components used above are those referring to the satellite-fixed PAI-system.

How shall the motion of the angular velocity vector be described in inertial space? It would of course be possible to take over the notation of eqn. (3.68). This would be awkward, however, because usually an equatorial (and not an ecliptical) coordinate system (referring to a standard epoch) is used to describe the motion of a satellite. This is why we introduce the angles Ω', i', and u' as Euler angles describing the angular velocity vector of the satellite. The analogy to the orbital motion is underlined in this definition of the Euler angles (Ω' corresponding to the right ascension Ω of the node of the orbital plane, i' corresponding to the inclination i of the orbital plane, and u' corresponding to the argument of latitude u of the satellite). Figure 3.10 illustrates the selected Euler angles of the satellite's attitude in the inertial, equatorial reference frame of a particular epoch. The figure should be compared to Figure 2.1. When adapting Euler's kinematic equations to the new set of angles given in Figure 3.10 we have to replace Ψ by Ω', $-\varepsilon$ by i', and Θ by u'. Equation (3.68), adapted to the new problem, thus reads as

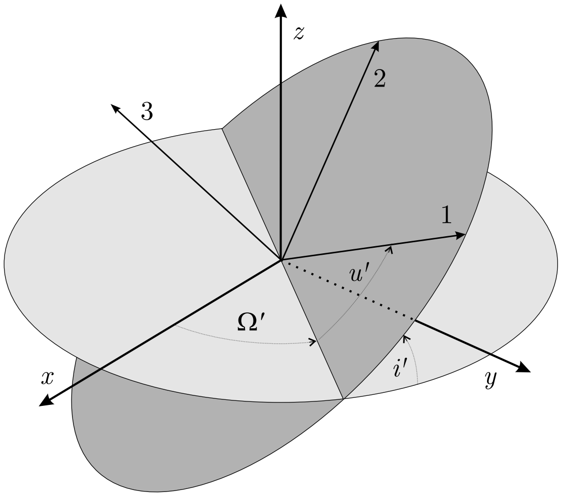


Fig. 3.10. Euler angles Ω', i', and u' describing the orientation of the satellite's PAI-system w.r.t. the equatorial system

$$\begin{pmatrix} \dot{\Omega}' \\ \dot{i}' \\ \dot{u}' \end{pmatrix} = \begin{pmatrix} +\sin u' \csc i' & +\cos u' \csc i' & 0 \\ \cos u' & -\sin u' & 0 \\ -\sin u' \cot i' & -\cos u' \cot i' & 1 \end{pmatrix} \boldsymbol{\omega}_s(t) \ . \tag{3.168}$$

Equations (3.167) and (3.168) completely describe the attitude of the satellite. On the right-hand side of eqns. (3.167) the coordinates of the satellite's center of mass in the satellite's PAI-system are required. The corresponding transformation matrix from the inertial to the satellite's PAI-system may be written as (see Figure 3.10)

$$\mathbf{T}_s^T = \mathbf{R}_3(u')\ \mathbf{R}_1(i')\ \mathbf{R}_3(\Omega') \ . \tag{3.169}$$

The transpose was used to preserve consistency of all transformation matrices $\mathbf{T}_{♁}$, $\mathbf{T}_m$, and $\mathbf{T}_s$ between the inertial and body-fixed systems. Assuming that the satellite was, at the time considered, at a geocentric distance r with an argument of latitude u, its coordinates in the satellite's PAI-system are

$$\begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \mathbf{R}_3(u')\ \mathbf{R}_1(i')\ \mathbf{R}_3(\Omega')\ \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-u) \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} \ . \tag{3.170}$$

Attitude Stabilization with the Gravitational Torque. Let us try to control the attitude of a satellite with the goal that

$$\begin{aligned} \delta\Omega &\overset{\text{def}}{=} \Omega' - \Omega \\ \delta i &\overset{\text{def}}{=} i' - i \\ \delta u &\overset{\text{def}}{=} u' - u \end{aligned} \tag{3.171}$$

remain small quantities. Initially, this may be achieved by an appropriate choice of an initial set of parameters. If eqns. (3.171) hold, the first coordinate axis of the PAI-system points approximately to the center of the Earth, the second along track (at least for circular orbits) and the third axis is normal to the (instantaneous) orbital plane. If we neglect all terms of higher than first order in these small quantities, the transformation equation (3.170) may be written as

$$\begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = r \begin{pmatrix} 1 \\ -\delta u \ - \ \cos i\ \delta\Omega \\ -\sin u\ \delta i \ + \ \sin i \cos u\ \delta\Omega \end{pmatrix} \ . \tag{3.172}$$

To the same order of approximation the Euler equations (3.167) may be written as

$$\begin{pmatrix} \dot{\omega}_{s_1} \\ \dot{\omega}_{s_2} \\ \dot{\omega}_{s_3} \end{pmatrix} + \begin{pmatrix} \gamma_{s_1}\, \omega_{s_2}\, \omega_{s_3} \\ \gamma_{s_2}\, \omega_{s_3}\, \omega_{s_1} \\ \gamma_{s_3}\, \omega_{s_1}\, \omega_{s_2} \end{pmatrix} = \frac{3\,GM}{r^3} \begin{pmatrix} 0 \\ \gamma_{s_2}\,(-\sin u\ \delta i \ + \ \sin i \cos u\ \delta\Omega) \\ -\,\gamma_{s_3}\,(\delta u \ + \ \cos i\ \delta\Omega) \end{pmatrix} \ . \tag{3.173}$$

Let us further simplify the problem by assuming rotational symmetry about the first axis, i.e., by putting $B_s = C_s$. This implies

$$\begin{aligned} \gamma_{s_1} &= \frac{C_s - B_s}{A_s} = 0 \\ \gamma_{s_2} &= \frac{A_s - C_s}{B_s} \overset{\text{def}}{=} \gamma_s \\ \gamma_{s_3} &= \frac{B_s - A_s}{B_s} = -\gamma_s \ . \end{aligned} \tag{3.174}$$

Equations (3.173) may then be modified as follows:

$$\begin{pmatrix} \dot{\omega}_{s_1} \\ \dot{\omega}_{s_2} \\ \dot{\omega}_{s_3} \end{pmatrix} + \gamma_s \begin{pmatrix} 0 \\ +\omega_{s_3}\,\omega_{s_1} \\ -\omega_{s_1}\,\omega_{s_2} \end{pmatrix} = \frac{3\,\gamma_s\,GM}{r^3} \begin{pmatrix} 0 \\ -\sin u\ \delta i \ + \ \sin i \cos u \ \delta\Omega \\ \delta u \ + \ \cos i \ \delta\Omega \end{pmatrix} \ . \tag{3.175}$$

As in the case of Earth rotation, the system is separated into one single (and trivial) equation, and into a system of two equations. In this case the constant rotation takes place about the first axis in the satellite's PAI-frame. This in turn transforms the system for the second and third component of the the angular velocity vector into a linear, inhomogeneous system. It might be solved in analogy to the solutions sketched in the case of Earth rotation.

We do not follow this procedure here, but reduce the problem further by approximating the orbit by an unperturbed elliptical orbit and by defining the initial state of the satellite's rotation as follows:

$$\begin{aligned} \delta i(t_0) &= 0 \quad &;& \quad \omega_{s_1}(t_0) = 0 \\ \delta\Omega(t_0) &= 0 \quad &;& \quad \omega_{s_2}(t_0) = 0 \\ \delta u(t_0) &= \delta u_0 \quad &;& \quad \omega_{s_3} = n_0 \ . \end{aligned} \tag{3.176}$$

where $n_0^2 = \frac{GM}{a^3}$ is the (Keplerian) mean motion of the satellite. With these assumptions and initial conditions, it is easy to verify that

$$\begin{aligned} \delta i(t) &= 0 \quad &;& \quad \delta\Omega(t) = 0 \\ \omega_{s_1}(t) &= 0 \quad &;& \quad \omega_{s_2}(t) = 0 \ . \end{aligned} \tag{3.177}$$

With this particular solution the third of Euler's kinematic equations (3.168) reads as:

$$\dot{u}_s = \dot{u} + \delta\dot{u} = n_0 + \int\limits_{t_0}^{t} \dot{\omega}_{s_3}(t')\ dt' \ . \tag{3.178}$$

This equation may be transformed into an ordinary differential equation by taking its first time derivative (and by rearranging its terms):

$$\delta\ddot{u} - \frac{3\,\gamma_s\,G\,M}{r^3}\;\delta u = -\,\ddot{u}\ . \tag{3.179}$$

Taking into account that, according to our assumptions, $r = r(t)$ and $\ddot{u} = \ddot{u}(t)$ are known functions of time t, the above equation is an ordinary, linear, non-homogeneous differential equation, which may be solved by standard procedures. Its structure is even better visible, if the problem is once more simplified by assuming that the orbit is circular. Then, the equation becomes even a homogeneous linear equation with constant coefficients

$$\delta\ddot{u} - 3\,\gamma_s\,n_0^2\;\delta u = 0\ . \tag{3.180}$$

The equation has periodic solutions, provided that

$$3\,\gamma_s\,n_0^2 < 0\ , \tag{3.181}$$

or, since n_0^2 is a positive quantity, provided that

$$C_s = B_s > A_s\ , \tag{3.182}$$

which means that the symmetric satellite must have one small and two large principal moments of inertia (a rod or an American football would meet the requirements) and the axis of minimum principal moment of inertia has to point (more or less) to the center of the Earth.

With the initial conditions specified above eqn. (3.180) has the solution

$$\delta u(t) = \delta u_0\,\cos(\tilde{\omega}_s(t - t_0))\ , \tag{3.183}$$

where

$$\tilde{\omega}_s = \sqrt{-\,3\,\gamma_s}\;n_0\ , \tag{3.184}$$

which means that the satellite's axis of minimum inertia oscillates about the radial direction with an amplitude defined (in one way or another) by the initial state of rotation. The oscillation period is given by

$$P_{\mathrm{att}} = \frac{1}{\sqrt{-\,3\,\gamma_s}}\;P_{\mathrm{rev}}\ , \tag{3.185}$$

where P_{rev} is the orbital period. The mechanism outlined is also called *gravity stabilization* of a satellite. It works under more general conditions than those considered here.

The best results (shortest period P_{att}) are obtained for $\frac{A_s}{B_s} \rightarrow 0$. $\frac{A_s}{B_s}$ is zero for an ideal rod with $A_s = 0$ and $B_s = C_s \neq 0$, where we have $\gamma_s = -1$, and the period P_{att} would be a factor of $\sqrt{3}$ shorter than the revolution period of the satellite. The period increases with increasing ratio $\frac{A_s}{B_s}$. For $A_s = B_s = C_s$

the period becomes infinite, indicating that a spherical satellite cannot be stabilized with this method. No periodic solutions exist for $A_s > B_s = C_s$. Gravity stabilization was, by the way, not invented by space agencies. The Moon and other natural satellites prove this statement.

We only addressed the principles of gravity stabilization here. For a more detailed discussion we refer to [103] and [41].

3.5 Relativistic Versions of the Equations of Motion

The equations of motion (3.13) of the planetary system and the corresponding equations (3.143) describing the motion of an artificial Earth satellite are approximations of the equations as they result from Einstein's general theory of relativity (or of even more general theories of gravitation), which were developed in the 20*th* century. A thorough treatment of this theory is outside the scope of this book, where the emphasis is on the *methods* of Celestial Mechanics. We refer to Soffel [109] for this purpose.

The relativistic version of these equations must be used, on the other hand, for many applications in practice, e.g., to generate the ephemerides of Sun, Moon, and planets as published in astronomical almanacs (see, e.g., [107], or [82]) and to compute precise (cm-type) orbits of artificial Earth satellites (see, e.g., [70]).

The corrections required by the general theory of relativity are very small. They may be taken into account by slightly modifying the equations of motion (3.13) of the planetary N-body problem and the corresponding equations of motion (3.143) of an artificial Earth satellite. The mathematical structure of the equations of motion is in essence preserved in this approximation. The equations of motion are quoted below (without proof) from the sources mentioned and their content is discussed. No attempt is made to describe the rotational motion of Earth and Moon in the framework of the theory of relativity.

When using the relativistic version of the equations of motion attention has to be paid to use the time argument which is consistent with the equations of motion. Terrestrial time (TT) is the correct time argument for the integration of the equations of motion (3.143) of satellite geodesy, TDB (Barycentric Dynamical Time) is the independent argument for the solution of the equations of motion of the planetary system (which are referred to the barycenter of the solar system). The approximate transformation between the two time scales is provided by eqn. (3.3), more precise formulae may be found in [107]. Both, TT and TDB are derived from the atomic time TAI.

According to [107] the relativistic version of the equations of motion of the planetary system, also called PPN (Parametrized Post-Newtonian) equations of motion of the planetary N-body problem read as

$$
\begin{aligned}
\ddot{\boldsymbol{x}}_i = & - k^2 \sum_{j=0,j\neq i}^{n} m_j \, \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \Bigg\{ 1 \; - \; \frac{2\,k^2\,(\beta+\gamma)}{c^2} \sum_{k=0,k\neq i}^{n} \frac{m_k}{|\boldsymbol{x}_i - \boldsymbol{x}_k|} \\
& - \frac{k^2\,(2\beta-1)}{c^2} \sum_{k=0,k\neq j}^{n} \frac{m_k}{|\boldsymbol{x}_j - \boldsymbol{x}_k|} \; + \; \gamma \, \frac{\dot{\boldsymbol{x}}_i^2}{c^2} \; + \; (1+\gamma)\, \frac{\dot{\boldsymbol{x}}_j^2}{c^2} \\
& - \frac{2\,(1+\gamma)}{c^2}\, \dot{\boldsymbol{x}}_i \cdot \dot{\boldsymbol{x}}_j \; - \; \frac{3}{2\,c^2} \left[\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)\; \dot{\boldsymbol{x}}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} \right]^2 - \; \frac{1}{2\,c^2}\, (\boldsymbol{x}_i - \boldsymbol{x}_j) \cdot \ddot{\boldsymbol{x}}_j \Bigg\} \\
& + \frac{k^2}{c^2} \sum_{j=0,j\neq i}^{n} \frac{m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \Big\{ (\boldsymbol{x}_i - \boldsymbol{x}_j) \big[\, (2 + 2\,\gamma)\, \dot{\boldsymbol{x}}_i \\
& - (1 + 2\,\gamma)\, \dot{\boldsymbol{x}}_j \,\big] \Big\} \cdot (\dot{\boldsymbol{x}}_i - \dot{\boldsymbol{x}}_j) \; + \; \frac{k^2\,(3 + 4\,\gamma)}{2\,c^2} \sum_{j=0,j\neq i}^{n} m_j \, \frac{\ddot{\boldsymbol{x}}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} \; , \\
& i = 0, 1, \ldots, n \; ,
\end{aligned}
\tag{3.186}
$$

where $c = 173.14463$ AU/d is the speed of light in AU per day, β and γ are parameters of the particular theory of gravitation used. For Einstein's general theory of relativity these parameters are

$$
\beta = \gamma = 1 \; . \tag{3.187}
$$

With the appropriate initial conditions and masses, eqns. (3.186) are the equations of motion describing the motion of the Sun, the planets, and (possibly) their moons. The bodies are considered to be point masses in an isotropic, PPN (Parametrized Post-Newtonian) N-body metric (see [109]).

The accelerations showing up on the right-hand side may be approximated by the non-relativistic equations (3.13).

Taking into account that

$$
\frac{k^2}{c^2} \approx 0.987 \cdot 10^{-8} \; , \tag{3.188}
$$

we see that the differences between the relativistic and the non-relativistic equations are of the order of (at maximum) a few parts in 10^{-8}.

Equations (3.186) are used to generate the ephemerides of Sun, Moon, planets and the asteroids Ceres, Pallas, Vesta, Iris, and Bamberga for the Astronomical Almanac. The complexity of eqns. (3.186) has the consequence that their numerical solution (as compared to that of eqns. (3.13)) is very inefficient. For applications over very long time spans (millions of years) the integration of eqns. (3.186) still is prohibitively slow – even when using modern computers.

For such applications it is useful to have a "light version" of eqns. (3.186) available. It is indeed possible to reduce these equations considerably by retaining only those correction terms proportional to m_0. This is justified under the assumption

$$m_i \ll m_0 \ , \quad i = 1, 2, \ldots, n \ , \tag{3.189}$$

i.e., for N-body problems describing a planetary system.

Assuming that eqn. (3.189) holds, one may in addition verify that $\dot{\boldsymbol{x}}_0$ is a small quantity. When calculating the correction terms it is therefore allowed to replace $\dot{\boldsymbol{x}}_i$ by the corresponding heliocentric velocity vector $\dot{\boldsymbol{r}}_i$. With these approximations one easily verifies that eqns. (3.186) for the bodies $i = 1, 2, \ldots, n$, (i.e., for the planets and moons) may be reduced to the following handy equations:

$$\ddot{\boldsymbol{x}}_i = -\,k^2 \sum_{j=0, j\neq i}^{n} m_j \, \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} + \frac{k^2\, m_0}{c^2\, r_i^3} \left\{ \left[4\, \frac{k^2\, m_0}{r_i} - \dot{\boldsymbol{r}}_i^2 \right] \boldsymbol{r}_i + 4\, (\boldsymbol{r}_i \cdot \dot{\boldsymbol{r}}_i)\, \dot{\boldsymbol{r}}_i \right\} . \tag{3.190}$$

The equations for the central body m_0 (index 0) do not contain sizeable correction terms in the sense mentioned above, which is why we may approximate them as

$$\ddot{\boldsymbol{x}}_0 = -\,k^2 \sum_{j=1}^{n} m_j \, \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \ , \tag{3.191}$$

which makes it easy to derive the relativistic equations for the relative (heliocentric) motion by taking the plain difference of eqns. (3.190) and (3.191)

Sometimes, the relativistic term is even further reduced. For *low eccentricity orbits* one may even argue that the scalar product

$$\boldsymbol{r}_i \cdot \dot{\boldsymbol{r}}_i \approx 0 \tag{3.192}$$

may be neglected. Moreover, the "energy theorem" of the two-body problem (see eqn. (4.20)) may be reduced for low eccentricity orbits to

$$\dot{\boldsymbol{r}}_i^2 \approx \frac{k^2\, m_0}{r_i} \ , \tag{3.193}$$

which allows it to reduce eqns. (3.190) to

$$\ddot{\boldsymbol{x}}_i \approx -k^2 \sum_{j=0, j\neq i}^{n} m_j \, \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \ + \ 3\, \frac{(k^2\, m_0)^2}{c^2} \, \frac{\boldsymbol{r}_i}{r_i^4} \ . \tag{3.194}$$

It is in essence this version of the relativistic equations of motion which was used in the long-term integration [95] of the planetary system. Observe, that eqns. (3.194) might be written as the gradient of a potential function, where the potential function would differ slightly from the $1/r$-potential.

When using program PLASYS (documented in Chapter II- 10) the equations of our solar system may be integrated in the form (3.18), corresponding to the classical Newton-Euler formulation, in the form (3.186) corresponding to the correct PPN-formulation (with the drawback that the integration becomes rather inefficient), or in the approximated version (3.190), which takes into account the corrections due to the general theory of relativity to within about 0.1%. When the latter option is selected, the integration is performed in the heliocentric system (using the corrections defined by eqns. (3.190)).

When solving the correct equations with program PLASYS the integration is performed in the barycenric system – which is why the equations for the Sun have to be integrated, as well. Let us mention, that in principle this might have been avoided by referring the equations (3.186) to the general relativistic definition of the center of mass (see, e.g., [107]):

$$\sum_{i=1}^{n} \tilde{m}_i \, \boldsymbol{x}_i = \mathbf{0} \ , \tag{3.195}$$

where

$$\tilde{m}_i = m_i \left\{ 1 \ + \ \frac{1}{2} \, \frac{\dot{\boldsymbol{x}}_i^2}{c^2} \ - \ \frac{1}{2\,c^2} \sum_{j=1, j \neq i}^{N} \frac{k^2 \, m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} \right\} \ . \tag{3.196}$$

Observe, that eqns. (3.195) are non linear in the coordinates $\boldsymbol{x}_i$.

The relativistic corrections required for an artificial Earth satellite are easily transcribed from eqns. (3.190) by replacing $k^2 \, m_0$ by GM, the product of the gravitational constant and the Earth's mass, and by using geocentric instead of heliocentric position vectors. The perturbing acceleration reads as (compare [70]):

$$\boldsymbol{a}_{rel} = \frac{GM}{c^2 \, r^3} \left\{ \left[4 \, \frac{GM}{r} - \dot{\boldsymbol{r}}^2 \right] \boldsymbol{r} \ + \ 4 \, (\boldsymbol{r} \cdot \dot{\boldsymbol{r}}) \, \dot{\boldsymbol{r}} \right\} \ , \tag{3.197}$$

where $c = 299792.458$ km/s is the speed of light, $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ are the satellite's geocentric position and velocity vector in a quasi-inertial system (one which does not rotate w.r.t. the inertial barycentric system).

Obviously the relativistic acceleration term $\boldsymbol{a}_{rel}$ lies in the orbital plane. For a close Earth satellite the acceleration due to the theory of relativity is of the order of 10^{-9} of the main term. For precise orbit determination it is mandatory to take such effects into account.

3.6 The Equations of Motion in Overview

Three sets of equations of motion were developed, namely

- the heliocentric equations of motion (3.18) for the planetary system in section 3.2;
- the geocentric equations of motion (3.118) for the centers of mass of Sun and Moon and the equations (3.124), (3.68) for the rotation of Earth and Moon in section 3.3;
- the geocentric equations of motion (3.143) for an artificial Earth satellite (and the equations (3.167) and (3.168) for the satellite's attitude in section 3.4).

These equations will be the basis for all our subsequent developments and considerations.

In all sets of equations of motion considered, the same pattern was used to derive the equations in their final form: Starting from (the modern understanding of) Newton's axioms the equations for each particle, be it a point mass or a mass element of an extended body, were set up in the inertial system. Then the equations were transformed to a body-centered system by subtracting the equations for the center of mass for the selected central body from the equations of all the other bodies. In the planetary system we obtained heliocentric, in the generalized three-body problem Earth-Moon-Sun and for artificial Earth satellites geocentric equations of motion.

Only the orbital motion had to be considered in section 3.2, equations for the rotation of bodies of finite extensions were derived as well in sections 3.3 and 3.4. The equations for the orbital and rotational motion are a consequence of the same mechanical principles, the same primitive (in the word's original sense) equations of motion for each particle (volume element) of a system. This pattern is perhaps best seen in the case of the three-body problem Earth-Moon-Sun, where the final equations for orbital and rotational motion all emerge from the same basic equations (3.81) by very simple operations: By integrating over the equations of all particles of a body, we obtained the equations of motion for the bodies' centers of mass, by first multiplying the basic equations (3.81) vectorially with their geo- or selenocentric position vector and then integrating over these equations the equations governing the rotation of Earth and Moon were obtained. That both, orbital and rotational motion, are a consequence of the same principles of mechanics was one of the deep insights Leonhard Euler gained in his article [34] around 1750.

We have seen furthermore, both for the system Earth-Moon-Sun and for the motion of an artificial satellite, that orbital and rotational motion are not independent. This implies that in principle the equations for the orbital and rotational motion should always be analyzed together. This is (almost) never done (and not required) in practice, because the coupling between the two types of motion is very weak.

An analysis of the equations of the three-body problem Earth-Moon-Sun showed that both, orbital motion and rotational motion, are governed by nonlinear differential equations in time. The structure of the rotational motion proved to be quite interesting, when analyzed in the body-fixed coordinate system. The equations for the time development of the angular velocity vector form a linear, inhomogeneous system of first order differential equations – allowing for approximate analytical solutions. In general, the equations for the angular velocity vector and for the Euler angles cannot be separated. The resulting system is a nonlinear first order system of ordinary differential equations for the components of the angular velocity vector and the three Euler angles.

An alternative treatment of this standard procedure was also given in section 3.3. Instead of using the components of the angular velocity vector and the three Euler angles to describe the rotation, it is possible to use the Euler angles and the components of the angular momentum vector. This alternative approach to solve the equations for the rotation is governed by eqns. (3.125) and (3.66).

The relativistic versions of the equations of motion (N-body problem and satellite motion) were included and discussed, but not derived. In all applications, which will be considered subsequently, it is perfectly allowed to treat the relativistic terms as small perturbations w.r.t. the classical equations of motion.

4. The Two- and the Three-Body Problems

The two-body problem is analyzed in section 4.1, the three-body problem in section 4.5. In both cases we assume point masses and neglect the effects due to the theory of general relativity.

4.1 The Two-Body Problem

4.1.1 Orbital Plane and Law of Areas

The *two-body problem* is governed by two point masses m_0 and m, where we call m_0 the primary, m the secondary mass (the index may be left out for the latter mass). The motion of m relative to m_0 results from equations (3.18) by retaining only the central mass m_0 and one of the other point masses. This implies that the perturbation term is zero and we obtain

$$\ddot{\boldsymbol{r}} = -\,k^2\,(m_0 + m)\,\frac{\boldsymbol{r}}{r^3} \overset{\text{def}}{=} -\,\mu\,\frac{\boldsymbol{r}}{r^3}\ , \tag{4.1}$$

where $\mu = k^2\,(m_0 + m)$.

We follow the procedure which led to the conservation law of total angular momentum in the N-body problem by multiplying eqn. (4.1) by the vector operator $\boldsymbol{r}\times$ and obtain

$$\boldsymbol{r} \times \ddot{\boldsymbol{r}} = -\,\mu\,\frac{\boldsymbol{r} \times \boldsymbol{r}}{r^3} \overset{\text{def}}{=} \boldsymbol{0}\ . \tag{4.2}$$

Obviously this implies (in analogy to the conservation of the angular momentum) that

$$\boldsymbol{r} \times \dot{\boldsymbol{r}} = \boldsymbol{h} \tag{4.3}$$

is constant in time. (Observe that the actual definition of vector $\boldsymbol{h}$ slightly differs from the angular momentum as defined by the equation (3.38)). Equation (4.3) implies that the motion of the point mass m relative to m_0 takes place in a plane through m_0. The orbital plane is defined by any set of position and velocity vectors $\boldsymbol{r}(t)$ and $\dot{\boldsymbol{r}}(t)$ (as long as the two vectors are not collinear).

Figure 4.1 shows that the vector $\boldsymbol{h}$ allows the computation of the orbital elements Ω, the longitude of the ascending node, and the inclination i, with respect to the fundamental plane of the inertial coordinate system (e.g., the ecliptic for applications in the planetary system):

$$\boldsymbol{h} = h \begin{pmatrix} \cos(\Omega - \frac{\pi}{2}) \sin i \\ \sin(\Omega - \frac{\pi}{2}) \sin i \\ \cos i \end{pmatrix} = h \begin{pmatrix} \sin \Omega \sin i \\ - \cos \Omega \sin i \\ \cos i \end{pmatrix} . \tag{4.4}$$

This allows the computation of the two orbital elements defining the orbital plane w.r.t. the fundamental plane chosen:

$$\Omega = \arctan \left(\frac{h_1}{-h_2} \right) , \qquad i = \arccos \left(\frac{h_3}{|\boldsymbol{h}|} \right) . \tag{4.5}$$

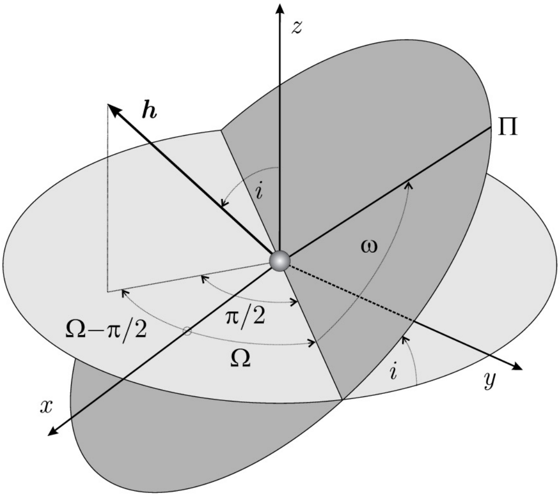


Fig. 4.1. Angular momentum vector $\boldsymbol{h}$ and orbital elements Ω and i

The length of the vector $\boldsymbol{h} = \boldsymbol{r} \times \dot{\boldsymbol{r}}$ is equal to the size of the area of the parallelogram spanned by the vectors $\boldsymbol{r}(t)$ and $\dot{\boldsymbol{r}}(t)$. The area dF swept up by the position vector $\boldsymbol{r}(t)$ during the infinitesimally short time interval dt is that of the triangle spanned by the vectors $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}\, dt$ (see Figure 4.2). Therefore, the area may be computed as:

$$dF = \tfrac{1}{2} \, |\boldsymbol{r} \times \dot{\boldsymbol{r}}| \, dt = \tfrac{1}{2} \, h \, dt \ , \tag{4.6}$$

where $h = |\boldsymbol{h}|$.

We recognize eqn. (4.6) as the infinitesimal formulation of Kepler's second law. By integration over time it emerges that the law holds for time intervals

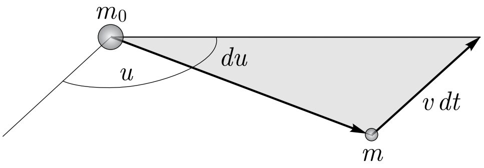


Fig. 4.2. Area filled by position vector $\boldsymbol{r}$ with velocity $\dot{\boldsymbol{r}}$ in short time interval dt

of arbitrary length. Kepler discovered this law for the elliptic motion in the planetary system. Kepler's second law for a time interval Δt is illustrated by Figure 2.2.

From its derivation we see that this law would hold for a broad class of force laws, not only for the inverse square law. In essence the right hand side of eqn. (4.1) must have the form $f(r)\,\boldsymbol{r}$. We also conclude from Figure 4.2 that we may write

$$dF = \tfrac{1}{2}\,r^2\,du \ , \tag{4.7}$$

where u and r are polar coordinates in the orbital plane. We identify u with the argument of latitude (see Figure 2.1), but this is not an important issue in our context; we might as well use another reference direction in the orbital plane.

By comparison of the right hand side of the above equation with that of eqn. (4.6) we obtain:

$$r^2\,du = h\,dt$$

or

$$\dot{u} = \frac{h}{r^2} \ . \tag{4.8}$$

4.1.2 Shape and Size of the Orbit

Following the procedure which led to the law of energy conservation in the N-body problem we multiply eqn. (4.1) by the vector operator $\dot{\boldsymbol{r}}\cdot$ and obtain

$$\dot{\boldsymbol{r}} \cdot \ddot{\boldsymbol{r}} = -\,\mu\,\frac{\dot{\boldsymbol{r}} \cdot \boldsymbol{r}}{r^3} \ .$$

Both sides of the equation may be written as time derivatives:

$$\frac{d}{dt}\left\{\tfrac{1}{2}\,\dot{\boldsymbol{r}}^2\right\} = \frac{d}{dt}\left\{\frac{\mu}{r}\right\} \ ,$$

which leads to the well-known result

$$\tfrac{1}{2}\,\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} = \tilde{E} \ , \tag{4.9}$$

where $\tilde{E}$ may be designated as the energy constant of the two-body problem. As $\dot{\boldsymbol{r}}^2$ and $\frac{\mu}{r}$ are both positive quantities, and as $\frac{\mu}{r} \rightarrow 0$ for $r \rightarrow \infty$, only for $\tilde{E} < 0$ the body m will never escape the main body. For $\tilde{E} = 0$ the body m has a velocity $|\dot{\boldsymbol{r}}| \rightarrow 0$ for $r \rightarrow \infty$, for $\tilde{E} > 0$ the absolute value of the velocity will be positive for $r \rightarrow \infty$. Equation (4.9) also implies that the absolute value of the velocity uniquely is a function of the distance between the two bodies for any given energy $\tilde{E}$.

Figure 4.2 illustrates that the velocity vector may be written as a superposition of two orthogonal unit vectors

$$\dot{\boldsymbol{r}} = \dot{r}\, \boldsymbol{e}_r \; + \; r\, \dot{u}\, \boldsymbol{e}_u \ , \tag{4.10}$$

where $\boldsymbol{e}_r$ is the unit vector in direction $\boldsymbol{r}$, $\boldsymbol{e}_u$ that in the orbital plane perpendicular to $\boldsymbol{r}$ (pointing into the direction of motion in the case of a circular motion). Squaring eqn. (4.10) leads to

$$\dot{\boldsymbol{r}}^2 = \dot{r}^2 + r^2\, \dot{u}^2 \ . \tag{4.11}$$

If we introduce formula (4.11) into equation (4.9) and use relation (4.8) to eliminate $\dot{u}$, we obtain a self-contained second order differential equation for the absolute value r of the position vector $\boldsymbol{r}$:

$$\frac{1}{2}\left(\dot{r}^2 + \frac{h^2}{r^2}\right) - \frac{\mu}{r} = \tilde{E} \ . \tag{4.12}$$

Equations (4.12) and (4.8) are transformations of the original equations of motion in the orbital plane. They separate our problem: We first solve eqn. (4.12) for $r(t)$ and introduce the result into eqn. (4.8) to obtain $u(t)$.

It is preferable to proceed in a slightly different way: Through the following transformation we replace the time t by the angular variable u

$$\dot{r} = \frac{dr}{du}\, \dot{u} = \frac{h}{r^2}\, \frac{dr}{du} \ .$$

This transformation allows us to study the shape of the orbital curve without considering the dynamics of the motion. Following the tradition (what would not be necessary) we also transform the dependent argument r by its inverse

$$\tilde{\sigma} \overset{\text{def}}{=} \frac{1}{r} \ . \tag{4.13}$$

We may then write:

$$\frac{dr}{du} = \frac{d}{du}\left(\frac{1}{\tilde{\sigma}}\right) = -\, \frac{1}{\tilde{\sigma}^2}\, \frac{d\tilde{\sigma}}{du} = -\, r^2\, \frac{d\tilde{\sigma}}{du} \ .$$

These transformations lead to the so-called *fundamental equation of the two-body problem*, also called *Clairaut's equation* in honour of Alexis-Claude Clairaut:

$$\left(\frac{d\tilde{\sigma}}{du}\right)^2 = \frac{2}{h^2}\left\{\tilde{E} + \mu\ \tilde{\sigma}\right\} - \tilde{\sigma}^2\ . \tag{4.14}$$

Clairaut's equation is a scalar, non-linear, first order differential equation for $\tilde{\sigma} = \frac{1}{r}$ with the argument of latitude u as independent argument. It is solved by

$$\tilde{\sigma} = \frac{1}{p}\left\{1 + e\cos(u - \omega)\right\}\ , \tag{4.15}$$

the polar equation for a conic section. p is the so-called *semi-latus rectum* of the conic section, e its numerical eccentricity, and ω is the argument of perihelion. This latter equation is better known when expressed in terms of the length r of the position vector $\boldsymbol{r}$ and the true anomaly $v = u - \omega$:

$$r = \frac{p}{1 + e\cos v}\ . \tag{4.16}$$

The square of the first derivative of expression (4.15) is

$$\left(\frac{d\tilde{\sigma}}{du}\right)^2 = \frac{e^2}{p^2}\sin^2(u-\omega) = \frac{e^2}{p^2} - \frac{e^2}{p^2}\cos^2(u-\omega) = \frac{e^2}{p^2} - \left\{\tilde{\sigma}^2 - \frac{2}{p}\tilde{\sigma} + \frac{1}{p^2}\right\}\ , \tag{4.17}$$

where we used eqn. (4.15) in the last step of the above derivation.

Comparing the coefficients of the terms $\tilde{\sigma}^i,\ i = 0, 1, (2)$, of this expression with those in Clairaut's equation (4.14), we verify that formula (4.15) solves the fundamental equation, provided the semi-latus rectum p and the eccentricity e of the conic section are defined as

$$p = \frac{h^2}{\mu} \quad \text{and} \quad e = \sqrt{1 + \frac{2\,h^2\,\tilde{E}}{\mu^2}}\ . \tag{4.18}$$

From the same formulae we may also compute the energy constant $\tilde{E}$ as a function of the semi-latus rectum of the ellipse and the eccentricity:

$$\tilde{E} = \frac{\mu}{2\,p}\left\{e^2 - 1\right\}\ . \tag{4.19}$$

The argument of perihelion is defined by the initial conditions associated with Clairaut's equation. So far, the solution of the two-body problem may be summarized as follows:

- Kepler's second law is a consequence of the conservation of angular momentum. The law implies that the motion is taking place in a plane defined by the orbital elements Ω and i (e.g., longitude of ascending node and inclination w.r.t. ecliptic).

- Energy conservation leads to Clairaut's equation, the fundamental equation of the two-body problem relating the absolute value r of the radius vector to the true anomaly v. Its solutions are conic sections (circle, ellipse, parabola, and hyperbola).
- The semi-latus rectum p, eccentricity e, and argument of perihelion ω are functions of h, the absolute value of the angular momentum vector $\boldsymbol{h}$, of the energy constant $\tilde{E}$, and of the initial conditions $\tilde{\sigma}(u_0)$ and $\frac{d\tilde{\sigma}}{du}(u_0)$ associated with Clairaut's equation. Kepler's first law thus is a consequence of energy conservation.
- So far, the solution of the two-body problem is defined by the five orbital elements p, e, i, Ω, and ω. These orbital elements are well suited to describe all conic sections.

The different orbit types are characterized in Table 4.1, where a is the semi-major axis of the conic section. The circle is a special case of the ellipse (with $e = 0$). From the orbit parameters p and e we may in particular calculate the pericenter distances (from m_0), and, in the case of the ellipse and the hyperbola, the semi-major axes. The semi-latus rectum p has the same simple geometric meaning for all conic sections: It is the length of the heliocentric position vector perpendicular to the major axis. Using these relations we may write the *energy theorem of the two-body problem* for the three cases as follows:

Table 4.1. Characterization of conic sections

Type	Eccentricity	Semi-latus rectum	Perihelion	Energy
Circle	$e = 0$	$p = a$	a	$-\frac{\mu}{2\,a} < 0$
Ellipse	$e < 1$	$p = a\,(1 - e^2)$	$a\,(1 - e)$	$-\frac{\mu}{2\,a} < 0$
Parabola	$e = 1$	p	$q = \frac{p}{2}$	$= 0$
Hyperbola	$e > 1$	$p = a\,(e^2 - 1)$	$a\,(e - 1)$	$+\frac{\mu}{2\,a} > 0$

$$\dot{\boldsymbol{r}}^2 = \mu \begin{cases} \left(\frac{2}{r} - \frac{1}{a}\right) , & \text{Ellipse} \\ \frac{2}{r} , & \text{Parabola} \\ \left(\frac{2}{r} + \frac{1}{a}\right) , & \text{Hyperbola} \end{cases} . \tag{4.20}$$

Note, that in some textbooks the semi-major axis a of the hyperbola is defined to be a negative value.

Figure 4.3 illustrates the three conic sections. The ellipses have an eccentricity of $e = 0.5$, the hyperbolas one of $e = 1.5$ (for the parabolas we have $e = 1$). The sun is at the coordinate origin. The orbits in the upper half-figure all have the same semi-latus rectum $p = 2.5$ AU, and in that crossing point even the same angular velocity (same distance $r = p$ from the Sun, same constant $h = \sqrt{\mu\, p}$; therefore the same $\dot{u} = \frac{h}{r^2} = \sqrt{\frac{\mu}{p^3}}$). The orbits in the lower half-figure have the same perihelion distance $\mid a\,(1-e) \mid = 2.5$ AU. Note also that the hyperbola asymptotes form angles of

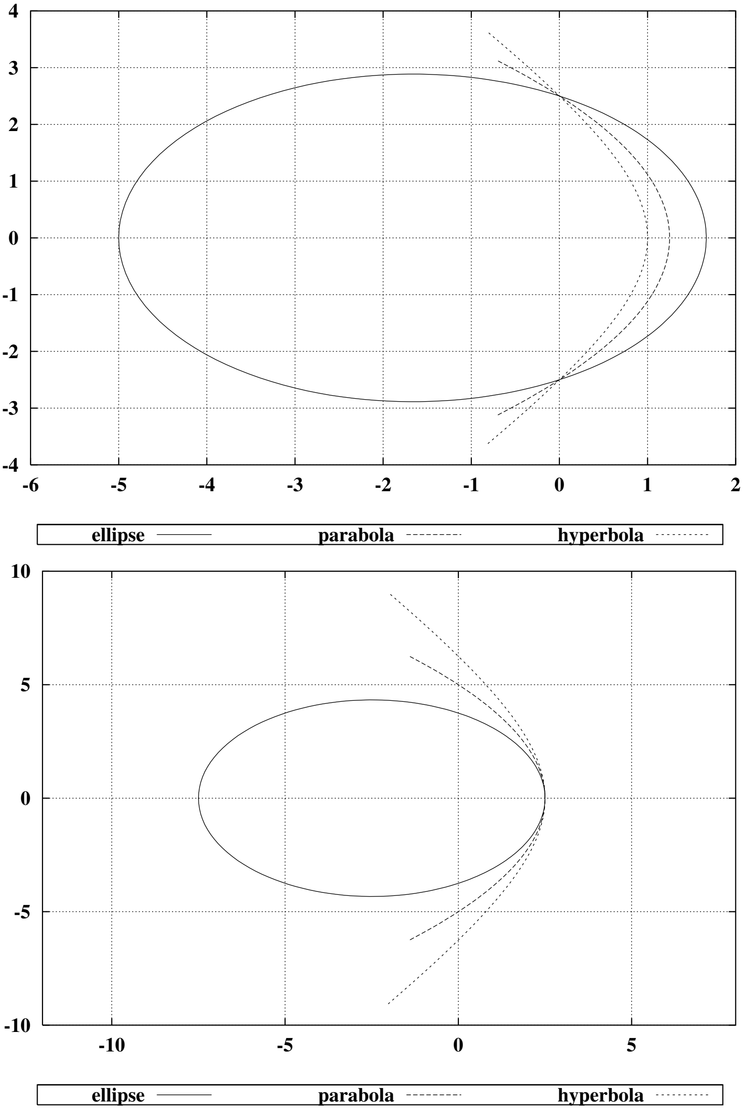


Fig. 4.3. Conic sections with the same semi-latus rectum p (upper) and the same perihelion (lower); eccentricities $e = 0.5, 1.0, 1.5$

$$\alpha = \pm \arctan \left\{ \sqrt{e^2 - 1} \right\} \ , \tag{4.21}$$

with the axis (sun-perihelion).

4.1.3 The Laplace Integral and the Laplace Vector $\boldsymbol{q}$

In mathematics it is not unusual that one and the same result may be obtained in different ways. In the previous section the orbital elements p and e were found by solving the fundamental equation (4.14) of the two-body problem. In this section we derive the so-called *Laplace vector* $\boldsymbol{q}$, a linear combination of the position and velocity vector, which is time-independent (i.e., a first integral). The Laplace vector may also be used to calculate the elements p and e. As frequent use will be made of the Laplace vector in the subsequent chapters, this second solution of (one part of) the two-body problem is a fruitful exercise.

Multiplying both sides of the equations of motion (4.1) of the two-body problem with the vector operator $\times \boldsymbol{h}$ (where $\boldsymbol{h}$ is the angular momentum vector (4.3) of the two-body problem) results in:

$$\ddot{\boldsymbol{r}} \times \boldsymbol{h} = -\frac{\mu}{r^3} \boldsymbol{r} \times \boldsymbol{h} \ . \tag{4.22}$$

Obviously the left-hand side of eqn. (4.22) may be written as a total time derivative:

$$\ddot{\boldsymbol{r}} \times \boldsymbol{h} = \frac{d}{dt} \left(\dot{\boldsymbol{r}} \times \boldsymbol{h} \right) \ . \tag{4.23}$$

Using the well-known relation

$$\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \cdot \boldsymbol{c}) \ \boldsymbol{b} - (\boldsymbol{a} \cdot \boldsymbol{b}) \ \boldsymbol{c} \tag{4.24}$$

from vector analysis, the right-hand side of eqn. (4.22) may be written as:

$$-\frac{\mu}{r^3} \boldsymbol{r} \times \boldsymbol{h} = -\frac{\mu}{r^3} \left\{ (\boldsymbol{r} \cdot \dot{\boldsymbol{r}}) \ \boldsymbol{r} - r^2 \ \dot{\boldsymbol{r}} \right\} = -\mu \left\{ \frac{\boldsymbol{r} \cdot \dot{\boldsymbol{r}}}{r^3} \boldsymbol{r} - \frac{\dot{\boldsymbol{r}}}{r} \right\} = \mu \frac{d}{dt} \left\{ \frac{\boldsymbol{r}}{r} \right\} \ . \tag{4.25}$$

The intermediary results(4.23) and (4.25) show that both sides of eqn. (4.22) may be written as total derivatives w.r.t. time t

$$\frac{d}{dt} \left\{ \dot{\boldsymbol{r}} \times \boldsymbol{h} - \mu \frac{\boldsymbol{r}}{r} \right\} = \boldsymbol{0} \ , \tag{4.26}$$

which is why, after integration, the following result may be established:

$$\dot{\boldsymbol{r}} \times \boldsymbol{h} - \mu \frac{\boldsymbol{r}}{r} = \boldsymbol{q} \ . \tag{4.27}$$

Equation (4.27) is referred to as the *Laplace integral*, the vector $\boldsymbol{q}$ as the *Laplace vector.* Applying formula (4.24) to the first term of eqn. (4.27) shows that the Laplace vector may be written as a linear combination of the position and velocity vector.

Even more insight into the structure of vector $\boldsymbol{q}$ is gained when multiplying eqn. (4.27) with the operator $\cdot\,\boldsymbol{r}$ (scalar product). The resulting scalar equation reads as:

$$(\dot{\boldsymbol{r}} \times \boldsymbol{h}) \cdot \boldsymbol{r} \;-\; \mu\, r = \boldsymbol{q} \cdot \boldsymbol{r} \overset{\text{def}}{=} q\, r\, \cos v \ , \tag{4.28}$$

where the angle between the position vector $\boldsymbol{r}$ and vector $\boldsymbol{q}$ was tentatively designated by v, the symbol reserved for the true anomaly v.

Using yet another result

$$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b} \cdot (\boldsymbol{c} \times \boldsymbol{a}) \tag{4.29}$$

from vector analysis, we see that v actually may be identified as the true anomaly: Using formula (4.29) to transform the first term in eqn. (4.28), this equation reads as

$$h^2 - \mu\, r = q\, r\, \cos v \ ,$$

which eventually leads to the polar equation of the conic sections

$$r = \frac{\frac{h^2}{\mu}}{1 + \frac{q}{\mu}\, \cos v} \ , \tag{4.30}$$

proving that the angle v actually *is* the true anomaly.

The result shows moreover that vector $\boldsymbol{q}$ is pointing to the pericenter and has the length μe, and that the semi-latus rectum p and the eccentricity e are defined by

$$p = \frac{h^2}{\mu} \qquad \text{and} \qquad e = \frac{q}{\mu} \ , \tag{4.31}$$

a result which was already established in eqn. (4.18) (the proof that the two results are algebraically identical is left to the reader).

It is often more convenient to use the vector

$$\boldsymbol{e} \overset{\text{def}}{=} \frac{\boldsymbol{q}}{\mu} \tag{4.32}$$

of length e instead of the Laplace vector $\boldsymbol{q}$. Both vectors, $\boldsymbol{q}$ and $\boldsymbol{e}$, will subsequently be referred to as Laplace vectors.

Let us conclude this section by explicitly writing the Laplace vector $\boldsymbol{q}$ as a linear combination of the position and velocity vector. Taking into account formula (4.24) eqn. (4.27) may be written as:

$$\left(\dot{\boldsymbol{r}}^2 - \frac{\mu}{r}\right) \boldsymbol{r} \;-\; (\boldsymbol{r} \cdot \dot{\boldsymbol{r}})\, \dot{\boldsymbol{r}} = \boldsymbol{q} \ . \tag{4.33}$$

Making use of the energy conservation law (4.20) this equation may be written in different ways for different orbit types. For elliptic orbits we have, e.g.:

$$\begin{aligned}\boldsymbol{q} &= \mu \left\{ \frac{1}{r} - \frac{1}{a} \right\} \boldsymbol{r} \;-\; (\boldsymbol{r} \cdot \dot{\boldsymbol{r}})\, \dot{\boldsymbol{r}} \\ &= \frac{1}{2} \left\{ \dot{\boldsymbol{r}}^2 - \frac{\mu}{a} \right\} \boldsymbol{r} \;-\; (\boldsymbol{r} \cdot \dot{\boldsymbol{r}})\, \dot{\boldsymbol{r}} \;.\end{aligned} \tag{4.34}$$

Observe, that $\boldsymbol{q} = \boldsymbol{0}$ for circular orbits.

4.1.4 True Anomaly *v* as a Function of Time: Conventional Approach

The transformation $t \rightarrow u$ of the independent argument removed the dynamics from the equations of motion and led to Clairaut's equation. If we introduce the solution (4.15) of this equation into the differential equation (4.8) for the argument of latitude u, the dynamics of the system is recovered by the following first order differential equation in the argument of latitude u:

$$\dot{u} = \frac{h}{r^2} = \frac{h}{p^2} \big(1 + e \cos{(u - \omega)}\,\big)^2 = \sqrt{\frac{\mu}{p^3}}\, \big(1 + e \cos{(u - \omega)}\,\big)^2 \;. \tag{4.35}$$

Its solution gives the argument of latitude u as a function of time. The equation may, e.g., be solved by the method of *separation of variables.* Using the *true anomaly* $v \overset{def}{=} u - \omega$ as angular argument (see also Figures 4.4 and 2.1), the solution reads as:

$$\int\limits_{u_0}^{u} \frac{du'}{\big(1 + e \cos(u' - \omega)\big)^2} = \int\limits_{v_0}^{v} \frac{dv'}{(1 + e \cos v')^2} = \sqrt{\frac{\mu}{p^3}}\; (t - t_0) \;\;, \tag{4.36}$$

where u_0 is the argument of latitude and v_0 the true anomaly at the initial epoch t_0.

u_0 would be the natural choice as sixth (and last) orbital element at time t_0. The set $p, e, i, \Omega, \omega, u_0$ of orbital elements describes the orbit of any two-body problem – for elliptic, parabolic, and hyperbolic orbits. Following astronomical tradition we modify eqn. (4.36) with the goal to start the integration in the perihelion. We replace the integration from v_0 to v by one from 0 to v and one from 0 to v_0:

$$\int\limits_{0}^{v} \frac{dv'}{(1 + e \cos v')^2} \;-\int\limits_{0}^{v_0} \frac{dv'}{(1 + e \cos v')^2} = \sqrt{\frac{\mu}{p^3}}\, \big((t - T_0) - (t_0 - T_0)\big) \;, \tag{4.37}$$

where T_0 is the time of pericenter passage. Instead of the "natural" orbit elements $p, e, i, \Omega, \omega, u_0$, one may therefore as well use the set $p, e, i, \Omega, \omega, T_0$.

If the orbit is either an ellipse or a hyperbola, p may be replaced by the semi-major axis a. We refer to Table 4.1 and to Figure 4.4 for the definition of the element a and the relationship between a, e and p in the case of the ellipse, to Figure 4.5 in the case of a hyperbola.

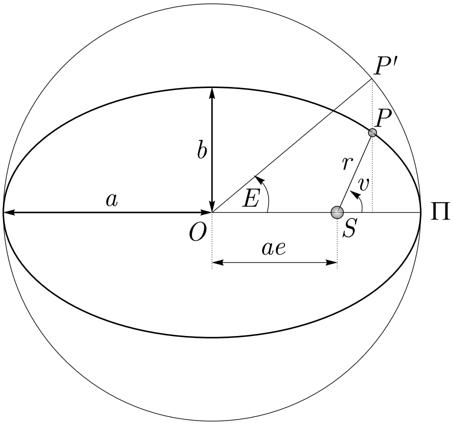


Fig. 4.4. True and eccentric anomalies v and E, semi-major axis a and eccentricity e in the elliptic orbit

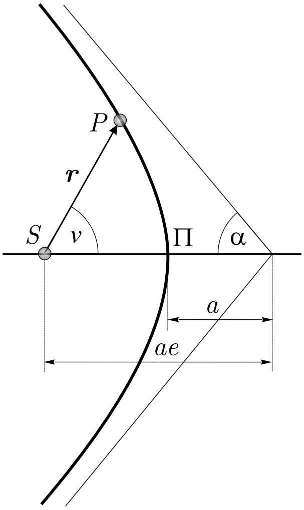


Fig. 4.5. True anomaly v, semi-major axis a and eccentricity e in the hyperbolic orbit

It is possible to compute the integral in eqn. (4.36) in closed form. This is only a technical issue, a matter of computational convenience and efficiency. It is only on this technical level where different cases, namely circles, ellipses, parabolas, and hyperbolas, have to be distinguished. Table 4.2 gives the conventional transformations required to solve equation (4.37), and the standard results.

Table 4.2. Equation for the true anomaly v: $\int\limits_0^v \frac{dv'}{(1+e\cos v')^2} = \sqrt{\frac{\mu}{p^3}}\ (t - T_0)$

Type	**Transformation**	**Solution**
Circle	—	$v = \sigma(t) \quad$ *)
Ellipse	$\tan\frac{v}{2} = \sqrt{\frac{1+e}{1-e}}\ \tan\frac{E}{2}$	$E - e\,\sin E = \sigma(t) \quad$ *)
Parabola	—	$\tan\frac{v}{2} + \frac{1}{3}\tan^3\frac{v}{2} = 2\sqrt{\frac{\mu}{p^3}}\ (t - T_0)$
Hyperbola	$\tan\frac{v}{2} = \sqrt{\frac{e+1}{e-1}}\ \tanh\frac{F}{2}$	$e\,\sinh F - F = \sigma(t) \quad$ *)

*) where: $\sigma(t) \overset{\text{def}}{=} n\ (t - T_0)$, with: $n \overset{\text{def}}{=} \sqrt{\frac{\mu}{a^3}}$

In Table 4.2 $\sigma = \sigma(t)$ is the *mean anomaly* for the elliptic and circular motion, the constant n is the mean motion. In the circular motion (special case $e = 0$) the true anomaly is the same as the mean anomaly $v(t) = \sigma(t)$. In the elliptic motion, the solution is sought by introducing the eccentric anomaly E (see Figure 4.4), which has a simple geometrical meaning. The solution is given by the *Kepler equation* relating the eccentric and mean anomalies E and σ.

Kepler's equation follows easily with the help of Figure 4.4 and Kepler's second law (law of areas): The area of the ellipse segment $SP\Pi$ is first computed with the help of the eccentric anomaly E and the elements a and e, and then with the law of areas:

$$\begin{aligned} A(SP\Pi) &= \frac{b}{a}\,A(SP'\Pi) = \frac{b}{a}\,\big(A(OP'\Pi) - A_\Delta(OP'S)\big) \\ &= \frac{b}{a}\,\frac{1}{2}\,\big(a^2\,E - a^2\,e\,\sin E\big) = \frac{ab}{2}\,(E - \sin E)\ . \end{aligned} \tag{4.38}$$

Kepler's second law implies on the other hand that

$$\frac{t - T_0}{U} = \frac{A(SP\Pi)}{ab\pi} \quad \text{or} \quad A(SP\Pi) = \frac{\pi}{U}\ \Delta t\,ab\ ,$$

where $\Delta t = t - T_0$ is the time elapsed between the planets perihelion passing and its arrival at position P, U is the revolution period, and $ab\pi$ is the area of the ellipse.

By equating the areas $A(SP\Pi)$ as computed in the previous two equations we obtain Kepler's equation:

$$E = \sigma(t) + e \sin E \ , \tag{4.39}$$

where $\sigma(t) \overset{\text{def}}{=} \frac{2\pi}{U}\,(t - T_0)$ is the mean anomaly of the celestial body at time t. This definition of the mean motion n (via the revolution time) is, by the way, which was accessible to Kepler. The relation $n \overset{\text{def}}{=} \sqrt{\frac{\mu}{a^3}}$ is only obtained by solving the equation (4.36) between the limits u_0 and $u_0 + 2\pi$ leading to the result:

$$\frac{2\pi}{\sqrt{(1-e^2)^3}} = \sqrt{\frac{\mu}{p^3}}\, U \ , \tag{4.40}$$

from where the relation

$$n^2\, a^3 = \mu \tag{4.41}$$

is easily obtained.

The Kepler equation is a linear equation for the determination of T_0 if E is given, it is transcendental in E, if σ is given. It has to be solved iteratively, where the simplest (and best known) algorithm reads as

$$\begin{aligned} E_1 \quad &\overset{\text{def}}{=} \sigma(t) = n\,(t - T_0) \\ E_{i+1} &= \sigma(t) + e\, \sin E_i \ , \quad i = 1, 2, 3, \ldots \ . \end{aligned} \tag{4.42}$$

This simple initialization is possible because $e < 1$ and $\sin i < 1$. For $e \ll 1$ the process converges rapidly, for higher eccentricities it is better to use an algorithm based on the correct linearization of the Kepler equation:

$$\begin{aligned} E_1 \qquad &\overset{\text{def}}{=} \sigma(t) = n\,(t - T_0) \\ \Delta E_{i+1} &= \frac{\sigma(t) - (E_i - e\, \sin E_i)}{1 - e\, \cos E_i} \\ E_{i+1} \quad &\overset{\text{def}}{=} E_i + \Delta E_{i+1} \ , \quad i = 1, 2, 3, \ldots \ . \end{aligned} \tag{4.43}$$

The analogue to Kepler's equation in the case of a hyperbolic orbit also is transcendental in F, the analogue to the eccentric anomaly E in the elliptic orbit. As opposed to the eccentric anomaly E, the angle F does not have a particular name, nor does it have a simple geometrical meaning. It is solved by a linearization of the analogue to Kepler's equation:

$$\begin{aligned} \Delta F_{i+1} &= \frac{-\sigma(t) + e\, \sinh F_i - F_i}{1 - e\, \cosh F_i} \\ F_{i+1} \quad &= F_i + \Delta F_{i+1} \ , \quad i = 1, 2, 3, \ldots \ . \end{aligned} \tag{4.44}$$

The initialization either may be done "graphically" or, e.g., by the following recipe found in [94]:

$$\begin{aligned} \sigma \le 5\,e - \tfrac{5}{2}: &\quad \text{solve} \quad F^3 + \frac{6\,(e-1)}{e}\,F - \frac{6\,\sigma}{e} = 0 \\ \sigma > 5\,e - \tfrac{5}{2}: &\quad \qquad\quad F = \ln\left(\frac{2\,\sigma}{e}\right)\ . \end{aligned} \tag{4.45}$$

The cubic equation in the above equations is of the structure

$$y^3 + a\,y = b\ , \tag{4.46}$$

which, according to [114], has exactly one real solution for $a > 0$:

$$y = \left\{\frac{b}{2} + \sqrt{\left(\frac{b}{2}\right)^2 + \left(\frac{a}{3}\right)^3}\right\}^{\frac{1}{3}} + \left\{\frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 + \left(\frac{a}{3}\right)^3}\right\}^{\frac{1}{3}}\ . \tag{4.47}$$

For the cubic equation in the recipe (4.45) we have

$$\frac{a}{3} = 2\,\frac{e-1}{e} \quad \text{and} \quad \frac{b}{2} = \frac{3\,\sigma}{e}\ , \tag{4.48}$$

giving rise to the solution

$$y = \left\{\frac{3\,\sigma}{e} + \sqrt{\left(\frac{3\,\sigma}{e}\right)^2 + \left(2\,\frac{e-1}{e}\right)^3}\right\}^{\frac{1}{3}} + \left\{\frac{3\,\sigma}{e} - \sqrt{\left(\frac{3\,\sigma}{e}\right)^2 + \left(2\,\frac{e-1}{e}\right)^3}\right\}^{\frac{1}{3}}. \tag{4.49}$$

The analogue to Kepler's equation for parabolic orbits is the polynomial equation of third degree in $\tan\frac{v}{2}$ given in Table 4.2. It is of the structure (4.46), which, in view of the above discussion, is solved by

$$\tan\frac{v}{2} = \left[\,\xi + \sqrt{\xi^2+1}\,\right]^{\frac{1}{3}} + \left[\,\xi - \sqrt{\xi^2+1}\,\right]^{\frac{1}{3}}, \quad \text{with} \quad \xi \overset{\text{def}}{=} 3\,\sqrt{\frac{\mu}{p^3}}\,(t-T_0)\ . \tag{4.50}$$

For elliptic orbits the eccentric anomaly E and for hyperbolic orbits the angle F proved to be most useful. When using these auxiliary quantities, it is important to provide the transformation between E and v, F and v, respectively, in a very explicit way. The transformation equations all may be derived from the the defining transformation provided in column 2 of Table 4.2 using the half-angle theorems of trigonometry (and the corresponding theorems for hyperbolic functions). The result is

$$\begin{aligned} \sin v &= \frac{\sqrt{1-e^2}\,\sin E}{1-e\,\cos E} \quad ; \quad \sin E = \frac{\sqrt{1-e^2}\,\sin v}{1+e\,\cos v} \\ \cos v &= \frac{\cos E - e}{1-e\,\cos E} \quad ; \quad \cos E = \frac{e+\cos v}{1+e\,\cos v}\ . \end{aligned} \tag{4.51}$$

The corresponding transformation equations for hyperbolic orbits read as

$$\begin{aligned} \sin v &= \frac{\sqrt{e^2-1}\,\sinh F}{e\cos F - 1} \quad ; \quad \sinh F = \frac{\sqrt{e^2-1}\,\sin v}{1+e\cos v} \\ \cos v &= \frac{e-\cosh F}{e\,\cosh F - 1} \quad ; \quad \cosh F = \frac{e+\cos v}{1+e\cos v} \ . \end{aligned} \tag{4.52}$$

Using the above transformation equations we may in particular immediately establish the relations for the absolute value r of the radius vector $\boldsymbol{r}$ for ellipses

$$r = \frac{p}{1+e\,\cos v} = a\,(1-e\,\cos E) \ , \tag{4.53}$$

and for parabolas:

$$r = \frac{p}{1+e\,\cos v} = a\,(e\,\cosh F - 1) \ . \tag{4.54}$$

4.1.5 True Anomaly v as a Function of Time: Alternative Approaches

The solution methods of the equation (4.36) for the true anomaly v as a function of time t outlined in the previous section were based on the introduction of (what might be called) auxiliary anomalies, namely the eccentric anomaly E for elliptic orbits and the corresponding quantity F for hyperbolic orbits. These anomalies are well established in astronomy and their use has important advantages. The iterative procedures for the computation of E (eqns. (4.43)) and F (eqns. (4.44)) are robust and converge rapidly.

On the other hand one may ask the question whether it is actually *necessary* to introduce case-specific anomalies? The question may be answered with a clear *no*: The true anomaly v has the same definition for the ellipse, the parabola, and the hyperbola; the true anomaly v is implicitly given by eqn. (4.36) together with some initial values, e.g., $v(T_0) = 0$; this initial value problem might then, e.g., be solved numerically using the technique of numerical quadrature (addressed in Chapter 7). The resulting algorithm of this "brute force" approach would, however, in general be very inefficient.

A much better alternative solution (without introducing auxiliary quantities) of eqn. (4.36) is possible by observing that the integral on the left-hand side may be solved analytically for all three cases.

The left interval boundary was chosen to correspond to the time of pericenter passage in the subsequent formulae. The corresponding integrals sometimes are referred to as the *flight-time equations* (e.g., [90]), because they allow it to compute the flight time since the (most recent) perihelion passage associated with a given value for the argument v. Using a standard compendium of mathematical formulae, e.g., [25], one may establish the relations:

$$\int\limits_0^v \frac{dv'}{(1+e\,\cos v')^2} = \sqrt{\frac{\mu}{p^3}}\;(t-T_0)$$

$$= \begin{cases} \frac{-\,e\sin v}{(1-e^2)(1+e\cos v)} + \frac{2}{\sqrt{(1-e^2)^3}}\;\arctan\left[\sqrt{\frac{1-e}{1+e}}\tan\frac{v}{2}\right] & ; \quad e<1 \\ \frac{1}{2}\left[\tan\frac{v}{2}+\frac{1}{3}\tan^3\frac{v}{2}\right] & ; \quad e=1 \\ \frac{e\sin v}{(e^2-1)(1+e\cos v)} - \frac{1}{\sqrt{(e^2-1)^3}}\;\ln\left[\frac{\sqrt{e^2-1}+(e-1)\tan\frac{v}{2}}{\sqrt{e^2-1}-(e-1)\tan\frac{v}{2}}\right] & ; \quad e>1\;. \end{cases} \tag{4.55}$$

The complexity of the first and the third of eqns. (4.55) may be slightly reduced by multiplying them with $\sqrt{(1-e^2)^3}$ and $\sqrt{(e^2-1)^3}$, respectively. Observing the definition of p in the case of the ellipse and the hyperbola (Table 4.1), one obtains:

$$\sqrt{\frac{\mu}{a^3}}\,(t-T_0) = \begin{cases} \dfrac{-\,e\sqrt{1-e^2}\,\sin v}{1+e\cos v} + 2\arctan\left[\sqrt{\dfrac{1-e}{1+e}}\,\tan\dfrac{v}{2}\right] & ; \quad e<1 \\ \dfrac{e\sqrt{e^2-1}\,\sin v}{1+e\cos v} - \ln\left[\dfrac{\sqrt{e^2-1}\,+\,(e-1)\,\tan\frac{v}{2}}{\sqrt{e^2-1}\,-\,(e-1)\,\tan\frac{v}{2}}\right] & ; \quad e>1\;. \end{cases} \tag{4.56}$$

Equations (4.55) and (4.56) are linear in the time t. As already mentioned above, it is therefore easily possible to calculate the epoch t corresponding to a given value of the true anomaly v.

The equations are transcendent in the true anomaly v. For a given epoch t it is a non-trivial task to compute the corresponding value for the true anomaly v – the exception being the second of eqns. (4.55) (which is identical with the equation already given in Table 4.2) and which is solved in closed form by eqn. (4.50).

The other two equations (for the ellipse and the hyperbola) may in principle be solved iteratively by some standard procedure of applied mathematics (e.g., a Newton-Raphson procedure).

Figure 4.6, showing the mean, eccentric, and true anomalies σ, E, and v for an orbit of eccentricity $e = 0.8$ as a function of the mean anomaly σ, illustrates that the convergence of the iteration process to determine v more critically depends on the first approximation than the corresponding iteration process for the determination of the eccentric anomaly E. According to Figure 4.6 the solutions for $\sigma = 40°$ are $E(\sigma) \approx 82°$ and $v(\sigma) \approx 141°$. Using the

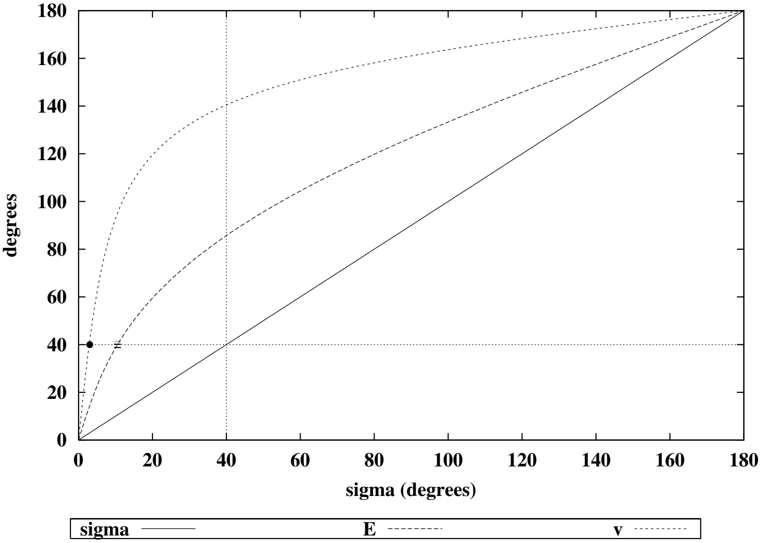


Fig. 4.6. Mean (σ), eccentric (E), and true anomalies (v) as a function of σ for $e = 0.8$

first approximation $E_1 \overset{\text{def}}{=} \sigma$ leads to a converging iteration process (4.50) for E (marked by a solid black square in Figure 4.6) whereas the analogue initialization $v_1 \overset{\text{def}}{=} \sigma$ (marked by a solid black circle in Figure 4.6) would lead to a diverging process. (This may be verified geometrically by constructing the tangents to the curves $E(\sigma)$ and $v(\sigma)$ at the marked points, intersect them with the vertical line at $\sigma = 40°$, find corresponding second approximation E_2 and v_2 by intersecting the horizontal line through the intersection points of the mentioned tangents and the vertical line at $\sigma = 40°$ with the curves $E(\sigma)$ and $v(\sigma)$, respectively).

The fact that the initialization of the iteration process for the determination of v is slightly more critical than for the corresponding processes in E and F does not invalidate the use of the alternative method. It just means that the iteration process has to be implemented carefully.

There are other alternatives to solve the integral for the anomaly. An attractive solution, which may be used for elliptic, parabolic, and hyperbolic orbits, consists in the introduction and computation of so-called universal variables. We refer to [31] for a concise discussion of the method.

4.2 State Vector and Orbital Elements

The position vector and the velocity vector together are also referred to as the *state vector* of the orbital motion. The state vector referring to one particular epoch and the corresponding differential equation system define one particular solution of the equations of motion.

The preceding analysis showed that the equations of motion of the two-body problem (4.1) are solved by conic sections as trajectories and that the time development is a solution of the equation (4.37) (or (4.36)) for the true anomaly v. Six time-independent orbital elements, e.g., the semi-latus rectum p of the conic section, the eccentricity e, the inclination i, the longitude of the ascending node Ω, the argument of perihelion ω, and the time of perihelion T_0 may be used to define the solution of the two-body problem. On the other hand, the state vector at epoch t also uniquely defines one particular solution of the equations of motion (4.1). Therefore there must be a one-to-one relationship between the orbital elements p, e, i, Ω, ω and T_0 and the state vector at any epoch t which will be formally established in the subsequent two paragraphs.

So far, all equations encountered in this Chapter were either vector equations *or* equations in the component matrices referring to the inertial system, which is characterized by the subscript $\mathcal{I}$. In order to reduce the formalism we left out the symbol $\mathcal{I}$, even if this would have been appropriate (like, e.g., in the equations (4.5), where the inclination i and the longitude Ω of the node were calculated as a function of the components of the angular momentum vector). We will preserve this habit throughout the book.

Subsequently we will need four different coordinate system in addition to the inertial system, all of them having a common fundamental plane, namely the orbital plane. We introduce the four systems for the two-body problem, not without pointing out, however, that the same coordinate systems may as well be defined and used for the perturbed motion. The difference merely resides in the fact that the orbital plane, which is fixed in the inertial space in the case of the two-body motion, has to be replaced by the instantaneous orbital plane referring to epoch t in the case of the perturbed motion.

As the fundamental plane is the orbital plane, these systems are called *orbital coordinate systems*. As the orbital plane is the common fundamental plane of all four systems, the systems also share the same third coordinate axis, which must be collinear with the angular momentum vector $\boldsymbol{h}$. Assuming all systems to be “Cartesian” (orthogonal, right-handed), the four different systems thus can thus be uniquely characterized by their first coordinate axes. Table 4.3 contains a list of the four orbital systems with their characterization, the definition of their first coordinate axis, and the coordinate transformation from the inertial system into the particular orbital system. Figure 4.7 illustrates the corresponding first axes.

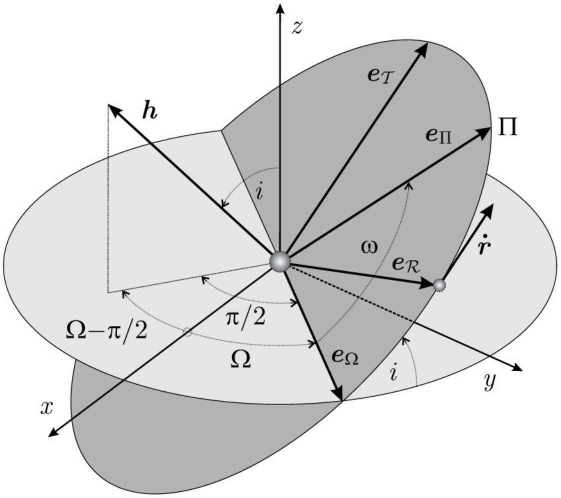


Fig. 4.7. First coordinate axes of orbital systems Ω, Π, $\mathcal{R}$, $\mathcal{T}$

Table 4.3. Orbital coordinate systems

System	First unit vector	Transformation from Inertial System $\mathcal{I}$	
Ω	$\boldsymbol{e}_\Omega = \frac{\boldsymbol{e}_3 \times \boldsymbol{h}}{h}$	$\boldsymbol{r}_\Omega =$	$\mathbf{R}_1(i)\ \mathbf{R}_3(\Omega)\ \boldsymbol{r}$
Π	$\boldsymbol{e}_\Pi = \frac{\boldsymbol{q}}{q}$	$\boldsymbol{r}_\Pi =$	$\mathbf{R}_3(\omega)\ \mathbf{R}_1(i)\ \mathbf{R}_3(\Omega)\ \boldsymbol{r}$
$\mathcal{R}$	$\boldsymbol{e}_\mathcal{R} = \frac{\boldsymbol{r}}{r}$	$\boldsymbol{r}_\mathcal{R} =$	$\mathbf{R}_3(u)\ \mathbf{R}_1(i)\ \mathbf{R}_3(\Omega)\ \boldsymbol{r}$
$\mathcal{T}$	$\boldsymbol{e}_\mathcal{T} = \frac{\dot{\boldsymbol{r}}}{\lvert\dot{\boldsymbol{r}}\rvert}$	$\boldsymbol{r}_\mathcal{T} =$	$\mathbf{R}_3(\xi)\ \mathbf{R}_3(\omega)\ \mathbf{R}_1(i)\ \mathbf{R}_3(\Omega)\ \boldsymbol{r}$

The coordinate transformations are given for the position vector in Table 4.3. The same transformation does of course hold for the velocity vector (or the acceleration vector).

Keep in mind that the argument of latitude u is defined by the sum of the argument of pericenter and the true anomaly $u = \omega + v$, and that the angle ξ is the angle between the Laplacian vector $\boldsymbol{q}$ (pointing to the pericenter) and the velocity vector $\dot{\boldsymbol{r}}$.

The transformation between the inertial system and each orbital system is defined by a series of rotations about particular axes and rotation angles. A particular rotation about an axis k and an angle α is defined by the transformation matrix $\mathbf{R}_k(\alpha)$, which is a 3×3 matrix. For practitioners we mention that the elements of the transformation matrix may be generated easily by an algorithm (easily translatable into a computer program) of the following kind:

1. $\mathrm{R}_{jj} \overset{\text{def}}{=} \cos\alpha \quad , \quad j = 1, 2, 3$
2. $\mathrm{R}_{12} = \mathrm{R}_{23} = \mathrm{R}_{31} \overset{\text{def}}{=} \sin\alpha$
3. $\mathrm{R}_{21} = \mathrm{R}_{32} = \mathrm{R}_{13} \overset{\text{def}}{=} -\sin\alpha$
4. $\mathrm{R}_{kj} = \mathrm{R}_{jk} \overset{\text{def}}{=} 0 \quad , \quad j = 1, 2, 3$
5. $\mathrm{R}_{kk} \overset{\text{def}}{=} 1 \quad .$

With this preparatory work it is now comparatively easy to solve our first task in this section, namely the calculation of the orbital elements from the state vector.

4.2.1 State Vector → Orbital Elements

Assuming that $\boldsymbol{r} \overset{\text{def}}{=} \boldsymbol{r}(t)$ and $\dot{\boldsymbol{r}} \overset{\text{def}}{=} \dot{\boldsymbol{r}}(t)$ are the components of the state vector in the inertial system at time t, we may determine the orbital elements as follows:

1. According to eqn. (4.3) we have
$$\boldsymbol{h} = \boldsymbol{r} \times \dot{\boldsymbol{r}} \ , \quad h = |\boldsymbol{h}| \ .$$
The orbital elements Ω and ω are defined by eqns. (4.5)
$$\Omega = \arctan\left(\frac{h_1}{-h_2}\right) \ , \qquad i = \arccos\left(\frac{h_3}{h}\right) \ .$$

2. The constant $\tilde{E}$ is obtained from equation (4.9)
$$\tilde{E} = \tfrac{1}{2}\,\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} \ ,$$
which allows the computation of the elements p and e with eqn. (4.18)
$$p = \frac{h^2}{\mu} \quad \text{and} \quad e = \sqrt{1 + \frac{2\,h^2\,\tilde{E}}{\mu^2}} \ .$$

3. Table 4.3 and Figure 4.7 show how to compute the argument of latitude at time t:
$$u \overset{\text{def}}{=} u(t) = \arctan\left(\frac{r_{\Omega_2}}{r_{\Omega_1}}\right) \ . \tag{4.57}$$

4. The coordinates of the position and velocity vector in the $\mathcal{R}$-System read as follows:
$$\boldsymbol{r}_{\mathcal{R}} = \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \dot{\boldsymbol{r}}_{\mathcal{R}} = \begin{pmatrix} \dot{r} \\ r\dot{u} \\ 0 \end{pmatrix} \ , \tag{4.58}$$

from where one may easily compute r and $\dot{u} = \dot{v}$, allowing it to retrieve the two orbital elements e and ω. Equation (4.15) and its first time derivative,

$$\dot{r} = -\frac{p}{\big(1 + e\,\cos(u-\omega)\big)^2}\,\big(-\,e\,\sin(u-\omega)\big)\dot{u} = \frac{r^2\,\dot{u}}{p}\,e\,\sin v\ , \qquad (4.59)$$

leads to the two equations for the determination of the true anomaly $v = u - \omega$:

$$e\,\cos v = \frac{p}{r} - 1\ ; \quad e\,\sin v = \frac{p}{r^2\,\dot{u}}\ , \qquad (4.60)$$

from where we may also determine the argument of perihelion

$$\omega = u - v\ . \qquad (4.61)$$

5. Using the equations in Table 4.2 (or the alternative formulae developed above) allows to determine the "last" element, the time T_0 of pericenter by solving, e.g., the flight-time equations (4.55) for the time argument $t - T_0$.

4.2.2 Orbital elements → State Vector

With the set p, e, i, Ω, ω, and T_0 of orbital elements we may derive the components of the vectors $\boldsymbol{r}(t)$ and $\dot{\boldsymbol{r}}(t)$ referring to the inertial system (or any other coordinate system we might choose) at any epoch t in the following steps:

1. The formulae in Table 4.2 (or the alternative formulae established) allow the computation of the true anomaly v.
2. For all orbit types the length r of the position vector $\boldsymbol{r}$ may now be computed using eqn. (4.16).
3. The components of the position vector at time t in the inertial coordinate system with the orbital plane as reference plane and the direction to the perihelion as first coordinate axis are defined as:

$$\boldsymbol{r}_\Pi = \begin{pmatrix} r\,\cos v \\ r\,\sin v \\ 0 \end{pmatrix} = \begin{pmatrix} a\,(\cos E - e) \\ a\,\sqrt{1-e^2}\,\sin E \\ 0 \end{pmatrix}\ . \qquad (4.62)$$

Figure 4.4 illustrates that in the case of the elliptic orbit the two components of position vector may be expressed either by the true anomaly v or the eccentric anomaly E. Note that the representation as a function of v holds for all types of conic sections. For the sake of completeness we note that the length r of the radius vector may also be expressed by the eccentric anomaly:

$$r = a\,(1 - e\,\cos E)\ . \qquad (4.63)$$

Equation (4.63) is equivalent to the polar equation (4.16).

4. The components of the velocity vector referring to the same coordinate system are derived by taking the first time derivative of eqn. (4.62):

$$\begin{aligned} \dot{\boldsymbol{r}}_{\Pi} = \begin{pmatrix} \dot{r}\,\cos v \,-\, r\,\sin v\,\dot{v} \\ \dot{r}\,\sin v \,+\, r\,\cos v\,\dot{v} \\ 0 \end{pmatrix} &= \sqrt{\frac{\mu}{p}} \begin{pmatrix} -\sin v \\ e + \cos v \\ 0 \end{pmatrix} \\ &= \dot{E} \begin{pmatrix} -\,a\,\sin E \\ a\sqrt{1-e^2}\,\cos E \\ 0 \end{pmatrix} , \end{aligned} \tag{4.64}$$

where we have used eqn. (4.8) with the understanding that $\dot{v} = \dot{u}$, because $u = v + \omega$, $\omega =$ const. The time derivative $\dot{E}$ is obtained by taking the time derivative of the Kepler equation (4.39). One easily establishes the result

$$\dot{E} = \frac{a\,n}{r} \ . \tag{4.65}$$

5. The components of vectors $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$ in the inertial system $\mathcal{I}$ are computed by the inverse sequence of rotations already given in Table 4.3:

$$\begin{aligned} \boldsymbol{r} &= \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \boldsymbol{r}_{\Pi} \\ \dot{\boldsymbol{r}} &= \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \dot{\boldsymbol{r}}_{\Pi} \ . \end{aligned} \tag{4.66}$$

6. If necessary, the components of the state vector may be transformed into other coordinate systems, e.g., into the geo- or topocentric equatorial systems.

We have thus demonstrated that for any epoch t there is a one-to-one relationship between the orbital elements and the state vector referring to epoch t:

$$\{\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)\} \leftrightarrow \{p, e, i, \Omega, \omega, T_0\} \ . \tag{4.67}$$

To the extent possible we have used formulae and parameters which are valid for all orbit types (ellipse, parabola, hyperbola).

4.3 Osculating and Mean Elements

In section 4.2 it was shown that the set of orbital elements defining the two-body problem may be computed from the state vector and that vice-versa the state vector referring to a particular epoch t may be computed from this unique set of orbital elements. In the framework of the two-body problem there is, in other words, a one-to-one correspondence between the orbital elements of a celestial body and the state vector referring to an (arbitrarily selected) epoch t:

$$t : \{\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)\} \leftrightarrow \{a, e, i, \Omega, \omega, T_0\} \ . \tag{4.68}$$

If the general N-body problem is integrated, the primary result consists of time series of state vectors for each of the constituents of the N-body system. If the orbit of a satellite is integrated in a very complex force field, the primary result consists of time series of geocentric state vectors for this satellite.

Using the formulae of the two-body problem it is possible to assign (for each celestial body considered) one set of orbital elements to each epoch t to the state vector (of the particular celestial body) of that epoch:

- Let $\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)$ be the solution of the equations of motion for one of the celestial bodies considered.
- The *osculating orbital elements* or *osculating elements* referring to the epoch t are defined by

$$t : \{\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)\} \rightarrow \{a(t), e(t), i(t), \Omega(t), \omega(t), T_0(t)\} \ , \tag{4.69}$$

  where t is called the *osculation epoch*, and where
- the osculating elements in the relation (4.69) are derived using the formulae of the two-body problem associated with the celestial body considered.

This concept of computing osculating elements is extensively used in programs PLASYS and SATORB, where the integration is performed in rectangular coordinates, but the planet-specific (satellite-specific) output contains time series of orbital elements (see Chapters II- 10 and II- 7 of Part III).

With the set of osculating elements referring to epoch t one may associate a Keplerian (two-body) orbit: The osculating orbit shares the state vector with the actual orbit at epoch t, but from there onwards (in positive and negative time direction) the osculating orbit follows the laws of the two-body problem (i.e., the osculating orbit curve is a conic section).

The actual orbit and the osculating Keplerian orbit are tangential at epoch t. This property explains the expression "osculating" stemming from the Latin verb "osculari", meaning "to kiss". The actual orbit is the envelope of all osculating orbits.

The osculating elements are excellent to gain insight of the orbital motion over few orbital periods. Figure 4.8 (left) illustrates the statement. The figure was generated with program PLASYS. The entire outer planetary system was integrated, the mentioned figure therefore illustrates the perturbations of the semi-major a of Jupiter due to the other planets over a few (about 10) revolutions. Figure 4.8 (right) illustrates, that the osculating elements are not ideal to study the development of an orbit over many revolutions (a time span of 2000 years corresponding to about 200 revolutions is covered in Figure 4.8 (right)). It would be preferable to remove the short period perturbations.

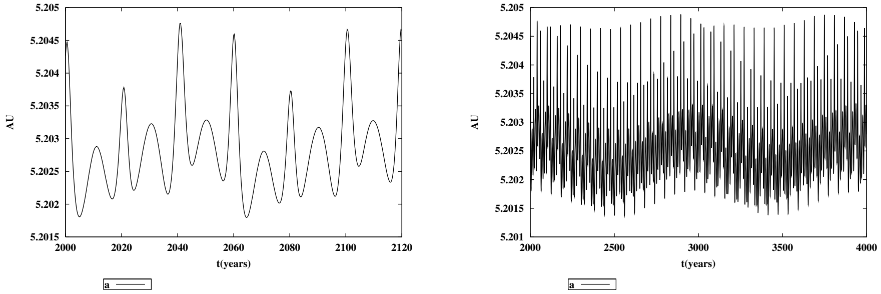


Fig. 4.8. Osculating orbital element a of Jupiter over a time intervals of 120 years (left) and 2000 years (right)

In Figure 4.8 (right) one would like to study other than short-period perturbations with relatively small amplitudes. In order to focus on these more interesting effects it is useful to define *mean orbital elements* in the following naive sense:

- Let
$$I(t) \in \{a(t), e(t), i(t), \Omega(t), \omega(t), T_0(t)\} \tag{4.70}$$
designate one of the osculating elements.
- The *mean orbital element* $\bar{I}(t; \Delta t(t))$, averaged over a time interval of $\Delta t(t)$ (which might be a function of time in the most general case), is defined as
$$\bar{I}(t; \Delta t(t)) = \frac{1}{\Delta t} \int\limits_{t-\Delta t/2}^{t+\Delta t/2} I(t')\, dt' \ . \tag{4.71}$$

Provided the osculating elements are continuous functions of time, eqns. (4.70) and (4.71) define a continuous function of time t, as well. If the averaging period Δt vastly exceeds all the short periods, or if the averaging period is an entire multiple of all short periods (something which is very difficult to achieve because there usually are no "strictly constant" periods in perturbed problems), the mean elements do no longer contain short period perturbations of significant size. Mean elements are therefore much better suited for studying the development of orbits over long time periods with thousands of revolutions.

The result of the averaging process (4.71) is illustrated in Figure 4.9 (right), where the osculating semi-major axis a of Jupiter was averaged over a time interval of five revolutions

$$\Delta t(t) \overset{\text{def}}{=} 5\, P_{♃}(t) \ , \tag{4.72}$$

$P_{♃}(t)$ being the (osculating) revolution period of Jupiter at time t. Note, that only by averaging over five (or an entire multiple of five) sidereal revolutions (corresponding to three synodical revolutions of the pair Jupiter-Saturn), (almost) all short period effects can be eliminated.

With the exception of calculating mean instead of osculating elements, Figure 4.9 (right) is based on an integration performed with identical options as that underlying Figure 4.9 (left), where the corresponding osculating element is shown. Obviously the removal of short period perturbations was rather successful. Extensive use of the concept of mean elements will be made in the subsequent chapters.

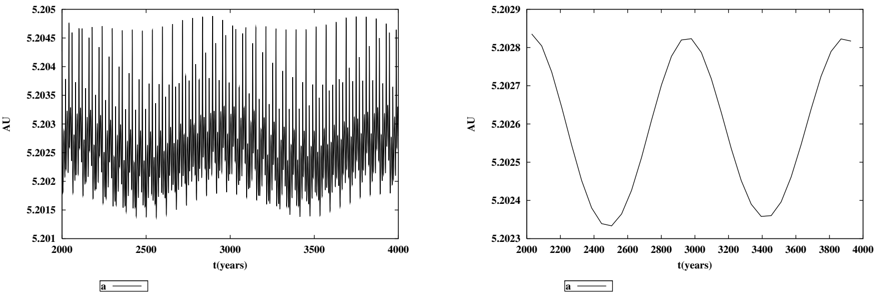


Fig. 4.9. Osculating (left) and mean (right) semi-major axis a of Jupiter over a time interval of 2000 years

The osculating and mean elements of Jupiter (and other planets) over time spans ranging from years to millions of years will be studied in more detail in Chapter II- 4.

4.4 The Relativistic Two-Body Problem

The equations for the relativistic motion of planets and satellites were introduced in section 3.5. From these equations we may extract the equations describing the relativistic two-body motion. For the subsequent treatment it is assumed that the conditions (3.189) hold, implying that we may use the simpler equations (3.190) (and not eqns. (3.186)) to take the relativistic effects into account.

Let us now consider the relativistic two-body problem with masses m_0 and m, assuming that $m \ll m_0$. The equation for the relative motion of the two bodies is obtained by taking the difference of eqns. (3.190), (3.191) resulting in

$$\ddot{\boldsymbol{r}} = -\,k^2\,(m_0 + m)\,\frac{\boldsymbol{r}}{r^3} \;+\; \boldsymbol{a}_{\text{rel}}\ , \tag{4.73}$$

with the perturbing acceleration

$$\boldsymbol{a}_{\text{rel}} = \frac{k^2\,m_0}{c^2\,r^3}\left\{\left[4\,\frac{k^2\,m_0}{r} - \dot{\boldsymbol{r}}^2\right]\boldsymbol{r} \;+\; 4\,(\boldsymbol{r}\cdot\dot{\boldsymbol{r}})\,\dot{\boldsymbol{r}}\right\}\ . \tag{4.74}$$

These equations have the same mathematical structure as eqns. (3.197) describing the perturbing accelerations for an artificial Earth satellite.

As viewed from classical (non-relativistic) theory the first term on the right-hand side of eqn. (4.73) is the two-body term of the classical theory, the second term is the “perturbation term” due to the theory of general relativity.

According to eqn. (4.74) the perturbation term $\boldsymbol{a}_{\text{rel}}$ is a linear combination of the relative position and velocity vector. The relativistic two-body motion therefore takes place in an orbital plane. The orbital elements Ω and i are thus first integrals of the relativistic motion, exactly as in the case of the non-relativistic two-body problem.

For low eccentricity orbits we may use the approximations

$$\boldsymbol{r}\cdot\dot{\boldsymbol{r}} \approx 0 \qquad \text{and} \qquad r \approx a = \text{const}\ ,$$

which allow it to reduce the differential equation (4.73) to

$$\ddot{\boldsymbol{r}} \approx -\,k^2\left(m_0 + m - \frac{3\,k^2\,m_0^2}{a\,c^2}\right)\frac{\boldsymbol{r}}{r^3} \overset{\text{def}}{=} -\,\tilde{\mu}\,\frac{\boldsymbol{r}}{r^3}\ . \tag{4.75}$$

Equations (4.75) are closely related to those of the classical two-body problem (4.1). The only difference resides in the fact that the product “gravity constant × mass of the central body” has to be slightly modified.

Our developments may be used to construct a circular orbit of radius a for the relativistic motion, as well. The “only” difference w.r.t. the classical circular orbit of radius a resides in the fact that the mean relativistic motion n_{rel} is defined by

$$n_{\text{rel}}^2\,a^3 = \tilde{\mu} = k^2\left(m_0 + m - \frac{3\,k^2\,m_0^2}{a\,c^2}\right)\ , \tag{4.76}$$

whereas the classical mean motion n is given by $n^2\,a^3 = k^2\,(m_0 + m)$. One easily verifies that

$$n_{\text{rel}} \approx n\,\sqrt{1 - \frac{3\,k^2\,m_0}{a\,c^2}} \approx n\left(1 - \frac{3\,k^2\,m_0}{2\,a\,c^2}\right)\ . \tag{4.77}$$

Using the value of $c = 173.14463$ AU/day for the speed of light, the numerical value (3.6) for the Gaussian constant k, and $m_0 = m_\odot$ for the solar mass one obtains the relativistic correction for the mean motion as

$$\delta n_{\mathrm{rel}} = 1.48 \cdot 10^{-8} \cdot \frac{n}{a} \ . \tag{4.78}$$

The impact δl due to the relativistic correction (4.78) in longitude l for a circular orbit with the semi-major axis of Mercury $a = 0.39$ AU after one century is then computed as:

$$\begin{aligned} \delta l &= \delta n_{\mathrm{rel}} \ \varDelta t \\ &= 4.046^\circ/\mathrm{day} \cdot 3.79 \cdot 10^{-8} \cdot 100 \cdot 365.25 \\ &= 0.0061^\circ/\mathrm{century} \\ &= 22.2''/\mathrm{century} \ , \end{aligned} \tag{4.79}$$

which is a very small value, indeed. It was virtually impossible to detect discrepancies of this kind before accurate distance measurements (e.g., Radar measurements to Venus in the 1960s) became available. A slightly wrong value for the mean motion of a planet could very well be absorbed by a slight change of the semi-major axis: Equation (4.41) allows it to calculate the relative error in the semi-major axis caused by a relative error in the mean motion and vice-versa. From

$$n^2 \, a^3 = k^2 \quad \text{one obtains:} \quad 2 \, n \, a^3 \, \delta n \ + \ 3 \, n^2 \, a^2 \, \delta a = 0 \ ,$$

and eventually

$$\frac{\delta n}{n} = -\frac{3}{2} \frac{\delta a}{a} \quad \text{and} \quad \frac{\delta a}{a} = -\frac{2}{3} \frac{\delta n}{n} \ . \tag{4.80}$$

Equation (4.78) tells that the (relative) relativistic correction of the mean motion is the order of a few parts in 10^8. Equations (4.80) tells that it is possible to absorb this effect by scaling (reducing) the semi-major axis by a factor $1 + \xi$, where ξ is a few parts in 10^8. As the semi-major axis of Mercury is $a \approx 0.39$ AU (corresponding to $a \approx 58500000$ km), the reduction to absorb the relativistic correction of the mean motion by the semi-major axis a is about 2 km – a discrepancy which was impossible to be detected, when the scale in the solar system was still established by the means of triangulation.

Mercury has an exceptionally large eccentricity of about $e \approx 0.206$ (see Table II- 4.1 in Chapter II- 4). The planet's perihelion therefore is well defined and can be observed accurately. Let us calculate the perturbations in the semi-major axis a, the eccentricity e and in the argument of perihelion ω using program PLASYS (see Chapter II- 10 of Part III). All other perturbations (due to the other planets) were "turned off". Figure 4.10 shows the resulting perturbations of the semi-major axis a and the eccentricity e. The Figure is based on the solution of the relativistic two-body problem Sun-Mercury (using the exact PPN-equations). The perturbations are shown only over the time interval of one year, because the perturbations are strictly periodic. The period of the perturbation is one revolution period of Mercury (about 0.24

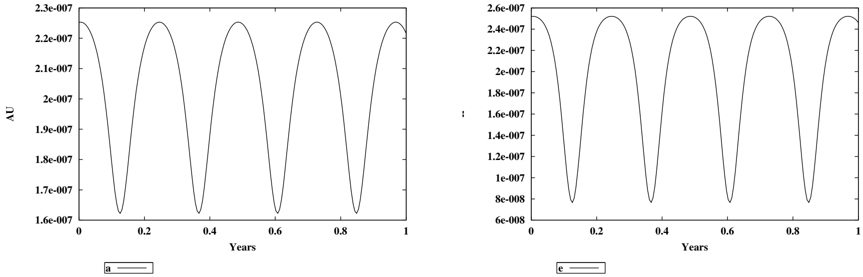


Fig. 4.10. Perturbation due to relativity in Mercury's semi-major axis a (left) and eccentricity e (right) over one year (two-body problem)

years), the amplitudes are of the order of $3 \cdot 10^{-8}$ AU in the semi-major axis, of the order of $9 \cdot 10^{8}$ in the eccentricity. Expressed in units of length, the variations are thus of the order of one kilometer. Unnecessary to say, that such subtle differences are extremely difficult to observe if only angles are measured.

Figure 4.11 show that Mercury's perihelion advances over one century as obtained by solving the relativistic two-body problem. As predicted by theory the advance is about $\Delta\omega = 43''$/century. This is a strong signal, which could be detected by optical observations already in the 19th century. Figure 4.12, giving Mercury's "actual" perihelion advance over 1000 years, calculated once with the Newton-Euler theory and once with the theory of relativity (the correct PPN-equations were used) including (in both integrations) all nine planets (Mercury-Pluto), demonstrates that the detection of Mercury's relativistic perihelion advance was far from trivial: Only about 4% of the total advance of about $1000''$/century are due to general relativity. Leverrier's merits (already mentioned in Chapter 2) are truly remarkable in this context.

4.5 The Three-Body Problem

After the successful treatment of the two-body problem the three-body problem is logically the next candidate of the N-body problem to be considered and solved. In the 18th century there was hope to find solutions in closed form, very much like they were described for the two-body problem. Investigations performed by Euler, Clairaut and Lagrange indicated, however, that simple analytical solutions would be obtained only under very special conditions. Today we know that the three-body problem is not solvable in analytically closed form and that it already contains most of the difficulties associated with the general N-body problem.

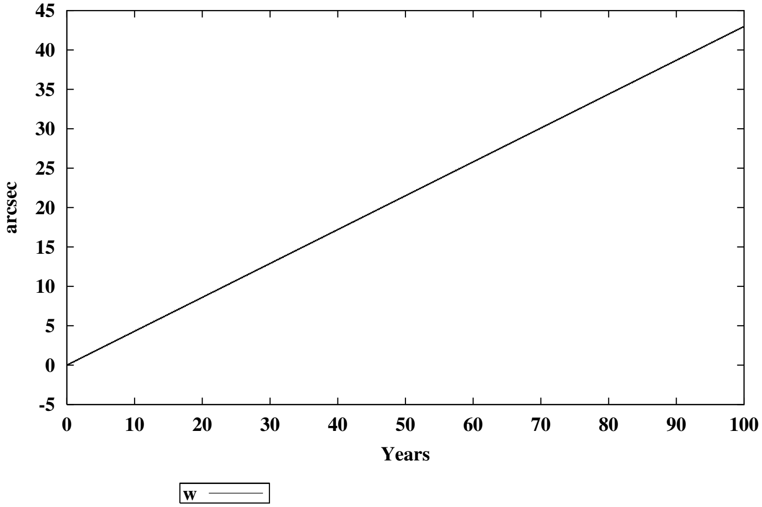


Fig. 4.11. Perturbation due to relativity in Mercury's argument of perihelion over 100 years (two-body problem)

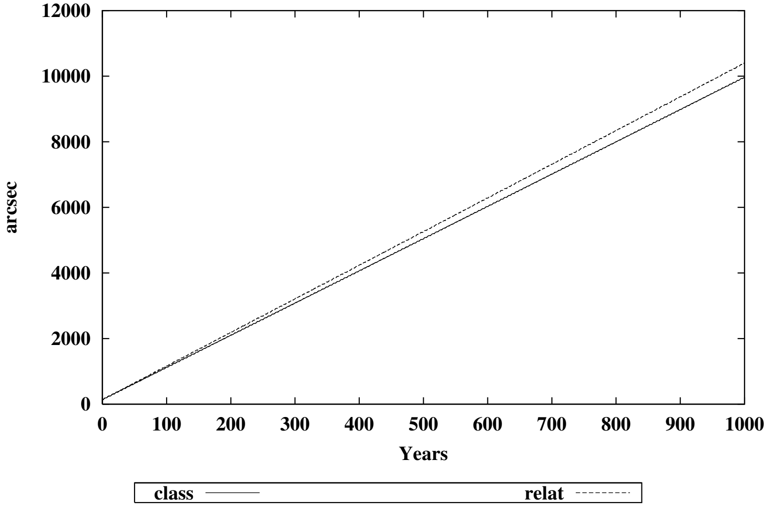


Fig. 4.12. Perturbation with and without relativistic correction in Mercury's argument of perihelion over 1000 years; all planets included

The three-body problem is relevant in practice. Think of the planetary system, which is governed by the three-body problem Sun-Jupiter-Saturn containing the greatest part of the mass of the solar system, of the problem Sun-Jupiter-asteroid, which is essential to study the long-term development of the belt of asteroids between Mars and Jupiter, the problems Earth-Sun-Satellite or Earth-Moon-spacecraft, which matter for space agencies wishing to fly to the Moon or to deploy Sun-observing satellites in the Earth-near space.

In the attempt to reduce the degree of difficulty of the problem, the so-called *problème restreint* or problema restrictum (in the scientific language of those days) was introduced by Euler. It was subsequently studied by many eminent mathematicians and astronomers. Euler, Lagrange, Carl Gustav Jacob Jacobi (1804–1851), Poincaré, Tissérand should be mentioned in particular. In the problème restreint it is assumed that the mass of one of the three bodies is small and may be neglected. This immediately reduces the three-body problem to describing the motion of the small body in the gravitational field of the two other bodies (revolving around each other in two-body orbits). In order to further reduce the degree of difficulty it is even assumed that the orbits of the two finite masses about each other are circular. For this version of the three-body problem Jacobi found a new integral of motion which is independent from the ten classical first integrals of the N-body problem (see section 3.2.2). This additional integral allows to gain considerable insight into the structure of the three-body problem. It will be derived and discussed in section 4.5.2. Despite all these positive aspects it must be admitted that closed analytical solutions could not even be found for the general case of the problème restreint. Poincaré even found chaotic aspects when studying the problème restreint as a first step to explore the stability of the solar system.

Some interesting properties of the general three-body problem will be dealt with in section 4.5.1, the problème restreint will be analyzed in section 4.5.2.

4.5.1 The General Problem

The general three-body problem with point masses is illustrated in Figure 4.13. We assume that the mass m_0 is the biggest of the three masses m_i , $i = 0, 1, 2$. The position vectors $\boldsymbol{r}_i$, $i = 1, 2$, are referred to m_0. S is the center of mass of the two bodies m_0 and m_1. Figure 4.13 also contains the *Jacobian vectors* $\boldsymbol{u}_i$, $i = 1, 2$, where $\boldsymbol{u}_1 \overset{\text{def}}{=} \boldsymbol{r}_1$ and where $\boldsymbol{u}_2 \overset{\text{def}}{=} \boldsymbol{r}_2 - \frac{m_1}{m_0+m_1}\,\boldsymbol{r}_1$ is the vector pointing from the center of mass S of m_0 and m_1 to the third body m_2. The ten classical integrals of the N-body problem were already established in section 3.2.2 and need not be repeated here. It is, however, instructive to express the conservation law for the angular momentum in the center of mass system

$$\boldsymbol{h} = m_0\,\boldsymbol{x}_0 \times \dot{\boldsymbol{x}}_0 \;+\; m_1\,\boldsymbol{x}_1 \times \dot{\boldsymbol{x}}_1 \;+\; m_2\,\boldsymbol{x}_2 \times \dot{\boldsymbol{x}}_2 \ , \tag{4.81}$$

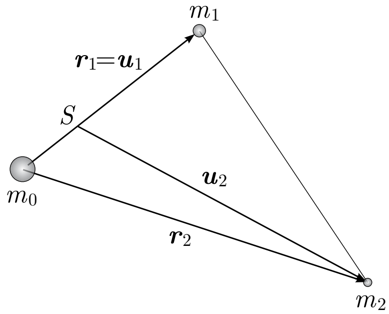


Fig. 4.13. The general three-body problem (point masses)

using the so-called Jacobian vectors (to be defined below). $\boldsymbol{x}_i$, $i = 0, 1, 2$, are the position vectors of the bodies with masses m_i in the inertial system (see section 3.2). After elementary algebraic transformations the following result is found:

$$\boldsymbol{h} = \frac{m_0\, m_1}{m_0 + m_1}\, \boldsymbol{u}_1 \times \dot{\boldsymbol{u}}_1 \; + \; \frac{m_2\,(m_0 + m_1)}{m_0 + m_1 + m_2}\, \boldsymbol{u}_2 \times \dot{\boldsymbol{u}}_2 \ . \tag{4.82}$$

The above relation is easily verified using the following definition for the Jacobian vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and the center of mass condition:

$$\begin{aligned} \boldsymbol{u}_1 \quad & \stackrel{\text{def}}{=} \boldsymbol{x}_1 \; - \; \boldsymbol{x}_0 \\ \boldsymbol{u}_2 \quad & \stackrel{\text{def}}{=} \boldsymbol{x}_2 \; - \; \frac{(m_0\, \boldsymbol{x}_0 + m_1\, \boldsymbol{x}_1)}{m_0 + m_1} \\ m_0\, \boldsymbol{x}_0 \; + \; m_1\, \boldsymbol{x}_1 \; + \; m_2\, \boldsymbol{x}_2 & = \boldsymbol{0} \ . \end{aligned} \tag{4.83}$$

Introducing the *fictitious masses* μ_1 and μ_2 by

$$\mu_1 \stackrel{\text{def}}{=} \frac{m_0\, m_1}{m_0 + m_1} \ , \quad \mu_2 \stackrel{\text{def}}{=} \frac{m_2\,(m_0 + m_1)}{m_0 + m_1 + m_2} \ , \tag{4.84}$$

the total angular momentum (4.82) may be composed as that of two subsystems, one consisting of the component μ_1 relative to the component m_0 and one consisting of the component μ_2 relative to the center of mass of the components m_0 and m_1:

$$\begin{aligned} \boldsymbol{h} & = \mu_1\ (\boldsymbol{u}_1 \times \dot{\boldsymbol{u}}_1) \; + \; \mu_2\ (\boldsymbol{u}_2 \times \dot{\boldsymbol{u}}_2) \\ & \stackrel{\text{def}}{=} \mu_1\, \boldsymbol{h}_1 \; + \; \mu_2\, \boldsymbol{h}_2 \ . \end{aligned} \tag{4.85}$$

If the mass m_0 dominates the two others, i.e., if $m_i \ll m_0$, $i = 1, 2$, we have approximately $\mu_1 \approx m_1$ and $\mu_2 \approx m_2$. These conditions are approximately met in the three-body problems of interest in the planetary system.

In the three-body problem Sun-Jupiter-Saturn, μ_1 would correspond approximately to the mass of Jupiter and $\boldsymbol{h}_1$ to the (two-body) angular momentum of Jupiter relative to the Sun; μ_2 would correspond to the mass of Saturn and $\boldsymbol{h}_2$ to the (two-body) angular momentum of Saturn relative to the Sun. In this approximation the angular momentum vectors $\boldsymbol{h}_i$, $i = 1, 2$, are perpendicular to the orbital planes of Jupiter and Saturn, respectively. This interpretation leads to Jacobi's theorem of nodes: Multiplying eqn. (4.85) with $\boldsymbol{h}\times$ to form a cross product one obtains the interesting relationship

$$\mu_1 \, \boldsymbol{h} \times \boldsymbol{h}_1 \approx - \mu_2 \, \boldsymbol{h} \times \boldsymbol{h}_2 \ , \tag{4.86}$$

where vector $\boldsymbol{h}$ is perpendicular to the so-called *invariable plane* (see section 3.2.2). Vectors $\boldsymbol{h} \times \boldsymbol{h}_i$, $i = 1, 2$, are thus vectors lying in the intersection of the invariable plane with the two orbital planes, thus in the line of nodes of the orbital planes (referring to the invariable plane as reference plane). More precisely the vectors point to the ascending nodes of the two planes. This explains the name *theorem of the nodes* associated with eqn. (4.86) for planetary three-body problems: Under the assumption $m_i \ll m_0$, $i = 1, 2$, the ascending node of Jupiter's orbit coincides with the descending node of Saturn's orbit. *Cum grano salis* we may therefore state that in a planetary three-body problem the orbital planes of the two small masses form approximately a rigid system, which may only rotate about the axis $\boldsymbol{h}$, the pole of the invariable plane.

Figure 4.14, which was generated with program PLASYS (documented in Chapter II- 10 of Part III), where only Jupiter, Saturn, and the Sun were included, showing the projection of the orbital poles (unit vectors perpendicular to the osculating orbital planes of Jupiter and Saturn on the plane of the ecliptic $J2000.0$), documents that Jacobi's theorem of the nodes is very well met by the three-body problem Sun-Jupiter-Saturn. The integration covers an interval of 40000 years. The lines connecting the starting and end points of the projections of the two orbital poles, respectively, intersect each other at the projection of the pole of the invariable plane on the ecliptic (see eqns. (II- 4.5) and (II- 4.9)). In Chapter II- 4 we will see that, even if the entire outer planetary system is included in the integration, Jacobi's theorem of the nodes still holds approximately.

The theorem of nodes is only *one* of the important aspects of the general three-body motion. Other aspects would be worth being discussed as well. Let us mention in particular the special cases which can be solved in closed form (first described by Euler and Lagrange). We will deal with them only under the more restrictive assumptions of the problème restreint in the next section. Moreover it would be attractive to study the intermediary lunar orbits of the three-body problem Earth-Moon-Sun. Space and time limitations do not allow it to discuss special problems of this kind here. For in-depth studies we refer to the standard treatment by Szebehely [117]. Also, Guthmann [49] offers a

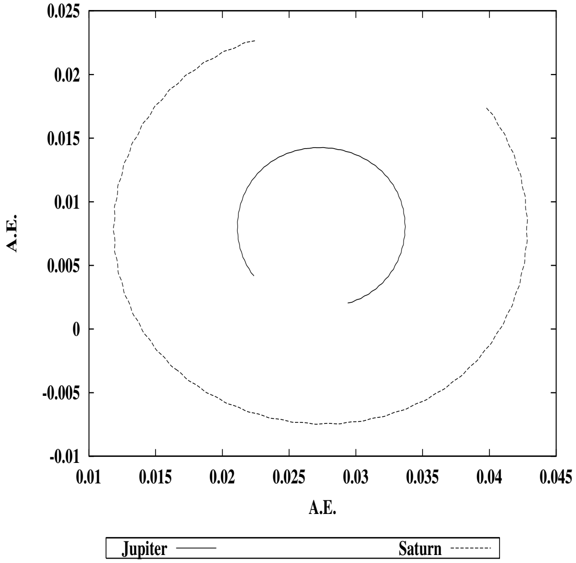


Fig. 4.14. Orbital poles of Jupiter's and Saturn's orbital planes seen from the North pole of the ecliptic $J2000.0$

concise treatment of the more important aspects of the general three-body problem in German. Danby [31] contains a concise treatment in English.

4.5.2 The Problème Restreint

In the problème restreint the motion of a point mass of negligible mass is studied in the gravitational field of two finite masses m_0 and m_1; the orbits of m_0 and m_1 are assumed to be circular. In his studies concerning the *problema restrictum* Euler even confined the discussion to a two-dimensional treatment by assuming that the third body would also move in the orbital plane of the bodies of finite mass.

Equations of Motion in the Inertial System. In any inertial system the motion of the three bodies is represented by (compare eqns. (3.13), section 3.2):

$$\ddot{\boldsymbol{x}} = -\,k^2 \left\{ m_0 \, \frac{\boldsymbol{x} - \boldsymbol{x}_0}{|\boldsymbol{x} - \boldsymbol{x}_0|^3} \; + \; m_1 \, \frac{\boldsymbol{x} - \boldsymbol{x}_1}{|\boldsymbol{x} - \boldsymbol{x}_1|^3} \right\} \; . \tag{4.87}$$

From now on it is assumed that eqns. (4.87) refer to the center of mass system of the bodies m_0 and m_1. It is natural to use the orbital plane of the two bodies as reference plane. Also, we assume that at epoch $t = 0$ the two bodies lie on the first (horizontal) axis of the coordinate system and that they rotate counterclockwise with angular velocity $\bar{n}$ (defined below). Their motion is illustrated by Figure 4.15. With these assumptions the motion of the two bodies (w.r.t. the rotating coordinate system of Figure 4.15) may be written as:

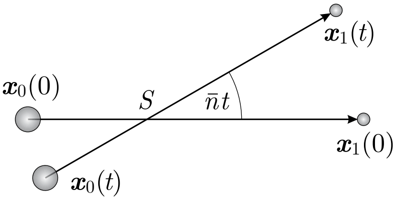


Fig. 4.15. Motion of the two bodies m_0 and m_1 in their orbital plane

$$\boldsymbol{x}_0 = -\frac{m_1}{m_0 + m_1}\,\bar{a}\begin{pmatrix}\cos\bar{n}t\\ \sin\bar{n}t\\ 0\end{pmatrix} \tag{4.88}$$

and

$$\boldsymbol{x}_1 = +\frac{m_0}{m_0 + m_1}\,\bar{a}\begin{pmatrix}\cos\bar{n}t\\ \sin\bar{n}t\\ 0\end{pmatrix}\;, \tag{4.89}$$

where $\bar{a} = |\boldsymbol{x}_1 - \boldsymbol{x}_0|$ is the radius of the circular motion of bodies m_0 and m_1 about each other; $\bar{n} = k\sqrt{\frac{m_0+m_1}{\bar{a}^3}}$ is the corresponding (mean) angular motion.

Equations in the Rotating Coordinate System. The equations of motion are now transformed into the coordinate system co-rotating with the two masses m_0 and m_1. The rotation axis is the pole of the orbital plane of the two finite bodies and the origin is the center of mass S. The first coordinate axis of the rotating system may be selected as the axis of the masses m_0 and m_1.

In the rotating system the coordinates of the celestial bodies are designated with $\boldsymbol{y}_0$, $\boldsymbol{y}_1$, and $\boldsymbol{y}$. The coordinates of m_0 and m_1 in the rotating system simply are:

$$\boldsymbol{y}_0 = \frac{m_1}{m_0 + m_1}\,\bar{a}\begin{pmatrix}-1\\ 0\\ 0\end{pmatrix} \tag{4.90}$$

and

$$\boldsymbol{y}_1 = \frac{m_0}{m_0 + m_1}\,\bar{a}\begin{pmatrix}1\\ 0\\ 0\end{pmatrix}\;. \tag{4.91}$$

For the test particle the transformation from the inertial into the rotating system reads as:

$$\boldsymbol{y} = \mathbf{R}_3(\bar{n}t)\, \boldsymbol{x} \ , \tag{4.92}$$

where $\mathbf{R}_3(\bar{n}t)$ is the 3 × 3-Matrix representing a rotation about the third coordinate axis by the angle of $+\bar{n}t$:

$$\mathbf{R}_3(\bar{n}t) = \begin{pmatrix} \cos\bar{n}t & \sin\bar{n}t & 0 \\ -\sin\bar{n}t & \cos\bar{n}t & 0 \\ 0 & 0 & 1 \end{pmatrix} \ . \tag{4.93}$$

The inverse transformation from the rotating into the inertial system for the test particle reads as:

$$\boldsymbol{x} = \mathbf{R}_3(-\bar{n}t)\ \boldsymbol{y} \ . \tag{4.94}$$

In order to transform the left-hand side of eqn. (4.87) into the rotating system, we have to calculate the first two time derivatives of eqns. (4.94):

$$\begin{aligned} \boldsymbol{x} &= \mathbf{R}_3(-\bar{n}t)\ \boldsymbol{y} \\ \dot{\boldsymbol{x}} &= \mathbf{R}_3(-\bar{n}t)\ \dot{\boldsymbol{y}} \ + \ \dot{\mathbf{R}}_3(-\bar{n}t)\ \boldsymbol{y} \\ \ddot{\boldsymbol{x}} &= \mathbf{R}_3(-\bar{n}t)\ \ddot{\boldsymbol{y}} \ + \ 2\,\dot{\mathbf{R}}_3(-\bar{n}t)\ \dot{\boldsymbol{y}} \ + \ \ddot{\mathbf{R}}_3(-\bar{n}t)\ \boldsymbol{y} \ . \end{aligned} \tag{4.95}$$

Substituting these relations into the equations of motion (4.87), multiplying the result with $\mathbf{R}_3(+nt)$, and using the matrix relations

$$\begin{aligned} \mathbf{R}_3(\bar{n}t)\ \mathbf{R}_3(-\bar{n}t) &= \mathbf{E} \ , \\ \mathbf{R}_3(\bar{n}t)\ \dot{\mathbf{R}}_3(-\bar{n}t) &= \bar{n} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \\ \mathbf{R}_3(\bar{n}t)\ \ddot{\mathbf{R}}_3(-\bar{n}t) &= -\,\bar{n}^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ , \end{aligned} \tag{4.96}$$

(where **E** is the identity matrix), one obtains the *equations of motion of the problème restreint* in the rotating coordinate system:

$$\ddot{\boldsymbol{y}} \ + \ 2\,\bar{n} \begin{pmatrix} -\dot{y}_2 \\ \dot{y}_1 \\ 0 \end{pmatrix} = \bar{n}^2 \begin{pmatrix} y_1 \\ y_2 \\ 0 \end{pmatrix} \ - \ k^2 \left\{ m_0\, \frac{\boldsymbol{y}-\boldsymbol{y_0}}{r_0^3} \ + \ m_1\, \frac{\boldsymbol{y}-\boldsymbol{y_1}}{r_1^3} \right\} \ , \tag{4.97}$$

where the symbols $r_0 \overset{\text{def}}{=} |\boldsymbol{y}-\boldsymbol{y}_0|$ and $r_1 \overset{\text{def}}{=} |\boldsymbol{y}-\boldsymbol{y}_1|$ were introduced. The second term on the left-hand and the first on the right-hand side, respectively, may be identified with the coriolis and centrifugal accelerations.

From the mathematical point of view the two systems of equations of motion (4.87) and (4.97) are equivalent. The latter allows us, however, to gain additional insight into the structure of the problem.

Jacobi's Integral. Multiplying the equations of motion (4.97) with $\dot{\boldsymbol{y}}$ one obtains:

$$\dot{\boldsymbol{y}}\,\ddot{\boldsymbol{y}} + \mathbf{0} = \bar{n}^2\,(y_1\,\dot{y}_1 + y_2\,\dot{y}_2) \;-\; k^2 \left\{ m_0\,\frac{\dot{\boldsymbol{y}}\,(\boldsymbol{y}-\boldsymbol{y_0})}{r_0^3} \;+\; m_1\,\frac{\dot{\boldsymbol{y}}\,(\boldsymbol{y}-\boldsymbol{y_1})}{r_1^3} \right\} . \tag{4.98}$$

Obviously, each term of the above equation represents a total time derivative (observe that vectors $\boldsymbol{y}_i\,,\ i=0,1$ are constant). Integrating the above equations and multiplying the result with the factor 2 we obtain the important formula:

$$\dot{\boldsymbol{y}}^2 = \bar{n}^2\left\{ y_1^2 + y_2^2 \right\} \;+\; 2\,k^2 \left\{ \frac{m_0}{r_0} + \frac{m_1}{r_1} \right\} \;-\; J\ . \tag{4.99}$$

Equation (4.99) represents a first integral which is *independent* of the classical ten integrals of the N-body problem. It is referred to as *Jacobi's integral* and the constant J occurring in it as *Jacobi's constant.*

Jacobi's integral is of central importance for the subsequent discussion of the problème restreint because it significantly constrains the motion of the test particle. It is, by the way, important to note that the terms on the right-hand side are invariant under the transformations (4.92, 4.94).

Tissérand-Criterion. Comets may have close encounters with Jupiter (or other planets). Their osculating orbital elements may have changed dramatically after such encounters. It is therefore in general not possible to decide whether two sets of osculating elements referring to epochs well before and after the epoch of a close encounter, respectively, belong to one and the same comet. We can expect, however, that the Jacobi constant J does not change by such a close encounter. The constant J thus is an ideal instrument to decide whether a "newly discovered" comet actually is identical with a known comet – rediscovered after a close encounter with Jupiter.

Tissérand gave the criterium a very useful form for practice. The corresponding criterium, although it is nothing but an application of Jacobi's integral, therefore is referred to as *Tissérand's criterion for the identification of comets.* Possibly, the monument erected in honour of Tissérand in Nuit St. Georges (Burgundy, France) (which contains, among other, a symbolized comet) has to be seen in the context of Tissérand's criterion.

Osculating elements are easily available for all known comets. The corresponding Jacobi constants usually are (were) not. Therefore, the question (at least at the times of Tissérand) whether there is a simple possibility to calculate Jacobi's constant using only the osculating elements is legitimate. A strikingly simple answer was given by Tissérand.

In order to derive this criterion we first write Jacobi's constant *approximately* as

$$J \approx \bar{n}^2 \left\{ y_1^2 + y_2^2 \right\} \, + \, 2\, k^2 \, \frac{m_0}{r_0} \, - \, \dot{\boldsymbol{y}}^2 \ , \tag{4.100}$$

where we assume that the above equation refers to an epoch well separated from that of a close encounter. Only under this condition one may neglect the term m_1/r_1 w.r.t. the term m_0/r_0 .

In order to express the right-hand side of eqn. (4.100) by the osculating elements, one has to transform it into the inertial system. As stated above the transformation merely consists of a rotation about the third coordinate axis about the angle $-\bar{n}t$. The length of the projection of a vector onto the reference plane is invariant under this transformation:

$$y_1^2 + y_2^2 = x_1^2 + x_2^2 \ . \tag{4.101}$$

The lengths r_0 and r_1 of the associated position vectors are of course invariant, as well. This only leaves us with the transformation of the term $\dot{\boldsymbol{y}}^2$. From the transformation equation (4.92) one may conclude:

$$\dot{\boldsymbol{y}}^2 = \Big(\dot{\boldsymbol{x}}^T \, \mathbf{R}_3(-\bar{n}t) \, + \, \boldsymbol{x}^T \, \dot{\mathbf{R}}_3(-\bar{n}t) \Big) \Big(\mathbf{R}_3(\bar{n}t) \, \dot{\boldsymbol{x}} \, + \, \dot{\mathbf{R}}_3(\bar{n}t) \, \boldsymbol{x} \Big) \ . \tag{4.102}$$

In analogy to the proof of relations (4.96) we easily verify that

$$\begin{aligned} \mathbf{R}_3(-\bar{n}t) \ \mathbf{R}_3(\bar{n}t) &= \mathbf{E} \ , \\ \dot{\mathbf{R}}_3(-\bar{n}t) \ \mathbf{R}_3(\bar{n}t) &= \bar{n} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \\ \dot{\mathbf{R}}_3(-\bar{n}t) \ \dot{\mathbf{R}}_3(\bar{n}t) &= \bar{n}^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ . \end{aligned} \tag{4.103}$$

Using the above relations in eqn. (4.102) and substituting the resulting expression into eqn. (4.100), we may calculate Jacobi's constant as:

$$J \approx 2\, k^2 \, \frac{m_0}{r_0} \, - \, \dot{\boldsymbol{x}}^2 \, + \, 2\, \bar{n} \, \left(x_1 \, \dot{x}_2 - x_2 \, \dot{x}_1 \right) \ . \tag{4.104}$$

Considering the fact that $x_1 \, \dot{x}_2 - x_2 \, \dot{x}_1$ is the third component of the angular velocity vector and making use of the energy theorem of the two-body motion, we obtain

$$J \approx \frac{k^2}{a} \, + \, 2\, k \, \bar{n} \, \sqrt{a \, (1 - e^2)} \, \cos i \ , \tag{4.105}$$

where we made use of the fact that in the solar system $m_0 = 1$.

Tissérand's criterion therefore may be stated as follows:

Two sets $\{t_j : \; a_j, \; e_j, \; i_j \, , \; j=1,2\}$ of osculating elements may be associated with one and the same comet, provided

$$\frac{k}{a_1} + 2\,\bar{n}\,\sqrt{a_1\left(1-e_1^2\right)}\,\cos i_1 \approx \frac{k}{a_2} + 2\,\bar{n}\,\sqrt{a_2\left(1-e_2^2\right)}\,\cos i_2 \; . \qquad (4.106)$$

Keep in mind that the above formula is only valid for epochs t_1 and t_2 well separated from an epoch corresponding to a close encounter. Observe, that in this approximation J is a strange combination of the energy and the third component of the angular momentum of the two-body problem. One might refer to it as an energy conservation law in the rotating system.

Hill's Surfaces of Zero Relative Velocity. For a given value of Jacobi's constant J Jacobi's integral (4.99) allows us to separate (in the rotating coordinate system) regions which are accessible to the test particle considered from those not accessible to it. The surfaces separating the regions are the *surfaces of zero relative velocity* (the attribute "relative" referring to a velocity in the rotating system). Requesting $\dot{\boldsymbol{y}}^2 \overset{\text{def}}{=} 0$ in the integral (4.99) defines a surface in the three dimensional space:

$$\bar{n}^2\left\{\,y_1^2+y_2^2\,\right\} + 2\,k^2\left\{\frac{m_0}{r_0}+\frac{m_1}{r_1}\right\} - J = 0 \; . \qquad (4.107)$$

Defining the reduced masses (masses relative to the total mass) by

$$m_0^* \overset{\text{def}}{=} \frac{m_0}{m_0+m_1} \, , \quad m_1^* \overset{\text{def}}{=} \frac{m_1}{m_0+m_1} \, , \qquad (4.108)$$

and the reduced, dimensionless Jacobi constant by

$$J^* \overset{\text{def}}{=} \frac{J}{\bar{n}^2\,\bar{a}^2} \, , \qquad (4.109)$$

the equations for the surfaces of zero relative velocity may be given in a slightly simpler form:

$$\frac{y_1^2+y_2^2}{\bar{a}^2} + 2\,\bar{a}\left\{\frac{m_0^*}{r_0}+\frac{m_1^*}{r_1}\right\} - J^* = 0 \; . \qquad (4.110)$$

The surfaces (4.110) are named after the American astronomer Hill. They separate those regions of the three-dimensional space for which $\dot{\boldsymbol{y}}^2 > 0$ (obviously that part of the space accessible to the body considered) from those for which $\dot{\boldsymbol{y}}^2 < 0$ (that part of space not accessible to the body).

It is instructive to draw the intersection of these surfaces with the coordinate planes for different values of the Jacobi constants. Such intersecting curves may be found in Figure 4.16 for a (hypothetical) mass ratio of $m_0 : m_1 = 3 : 1$

(Figures on left-hand side) and for a mass ratio of $m_0 : m_1 \approx 1047 : 1$ (the mass ratio of the solar mass and Jupiter's mass) on the right-hand side. The first row of Figures shows the intersection of Hill's surfaces with the (x, y)-plane, the second row the intersection with the (x, z)-plane, and the third the intersection with the (y, z)-plane. The zero-velocity curves in the left column are drawn for the values $J^* = 4.0, 3.8, 3.2, 2.8$, of the reduced Jacobi constant, those in the right column for $J^* = 4.0, 3.5, 3.0, 2.8$. In the Figures 4.16 (left) the projections of the two masses m_0 and m_1 are included as "+". In the projection onto the (x, y)-plane (top, left) the five stationary solutions $L_1 - L_5$ are marked with "x". These solutions will be discussed below.

Equation (4.110) says that the reduced Jacobi constant J^* (thus also the constant J) always must be positive. For big values of J^* a positive value for the square of the velocity $\dot{\boldsymbol{y}}^2$ only results *either* in the immediate vicinity of the two masses m_0 and m_1 *or* for big values of $y_1^2 + y_2^2$, i.e., if the test particle is far away from the z-axis. This in turn implies, that the test particle is either trapped in an (almost) spherical vicinity of the masses m_0 and m_1 or it must be outside of a cylindrically shaped boundary of about $7 - 8$ AU. The innermost boundary surfaces around m_0 and m_1 and the outermost cylindrical boundary in Figure 4.16 correspond to the maximum value of $J^* = 4.0$ of the Jacobi constant.

If the Jacobi constant J^* decreases, the permissible regions around the celestial bodies become bigger and bigger, and they are deformed to become a connected dumb-bell shaped region. This occurs between $J^* = 4.0$ and $J^* = 3.8$ in the left column of Figures 4.16. In addition, the radius of the boundary cylinder becomes smaller and smaller in the vicinity of the (x, y)-plane.

If the Jacobi constant is further reduced, the boundary cylinder and the dumb-bell shaped regions are connected. For $J^* = 3.2$ two kidney-shaped regions are left back in the (x, y)-plane. For the value $J^* = 2.8$ of the Jacobi constant the entire (x, y)-plane is accessible to the test particle. The zones of avoidance are contained within cylindrically shaped surfaces – moving further and further away from the (x, y)-plane (see left column, middle and bottom row of Figures 4.16).

Usually, Hill's surfaces of zero relative velocity are drawn for mass ratios of the type $(m_0 : m_1) = (2 : 1), (3 : 1), (3 : 2), \ldots$. From the "designer's" point of view one obtains the nicest figures for values of this kind. The right column of Figures 4.16, which was drawn for the mass ratio $(m_\odot : m_♃) \approx 1047 : 1$, demonstrates that more realistic scenarios in the planetary system result in slightly less attractive figures. For the value $J^* = 4, 3.5$ the allowed region around the mass m_{jup} are very small indeed (they are only visible as points in the top and middle row of Figures 4.16). For the values of $J^* = 3.0$ the two kidney-shaped areas are actually connected (Figures 4.16, top, left).

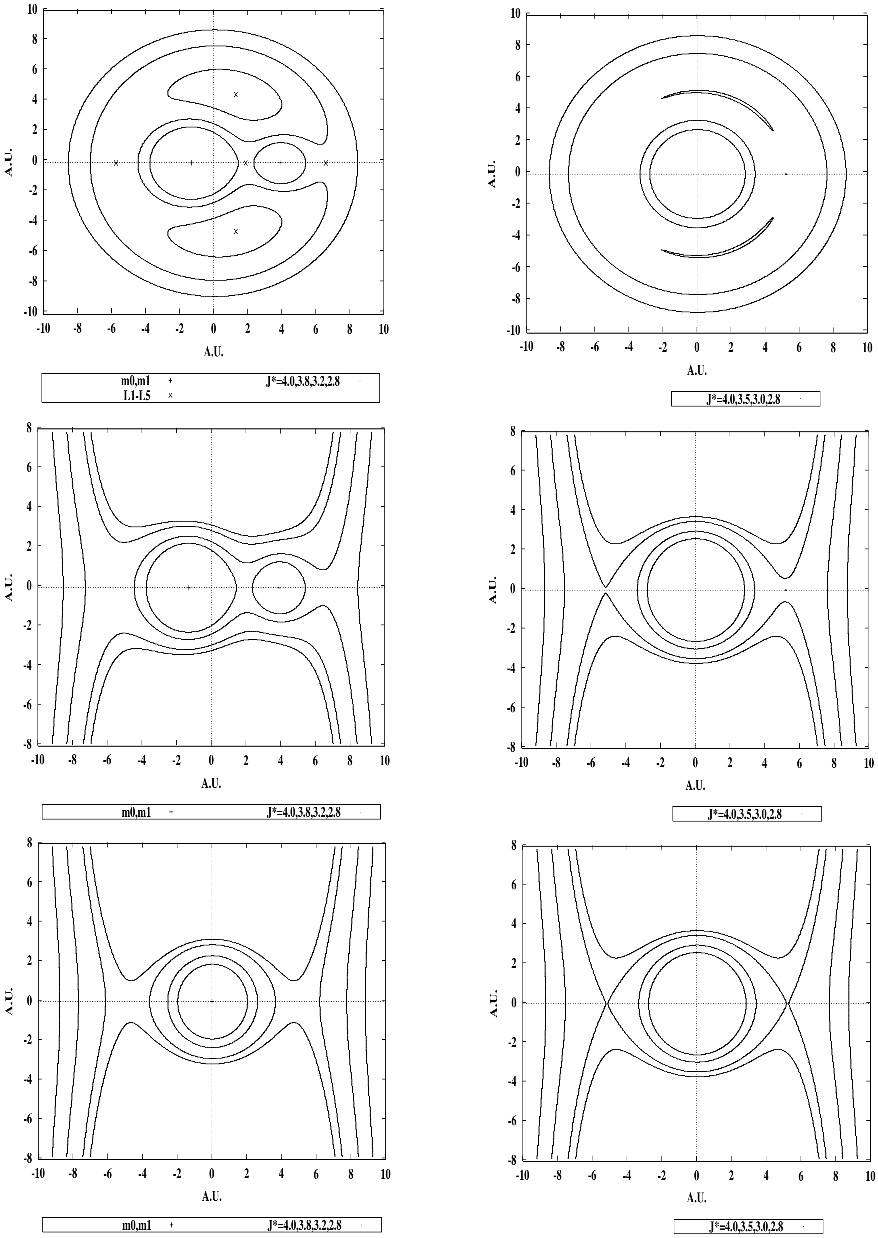


Fig. 4.16. Hill's surfaces ((y_1, y_2)-plane top, (y_1, y_3)-plane middle, (y_2, y_3)-plane bottom) of zero velocity for $(m_0 : m_1) = (3 : 1)$ (left) and for the case Sun-Jupiter (right)

Stationary Solutions. In this paragraph we are looking for solutions of the equations (4.97) which are at rest in the rotating system. We call them *stationary solutions* of the problème restreint. For a stationary solution we must have for arbitrary epochs t:

$$\boldsymbol{y}(t) = \boldsymbol{z}_0 = \text{const.} \tag{4.111}$$

Such solutions obviously result if, at a particular initial epoch t_0, the following conditions are met:

$$\begin{aligned} \boldsymbol{y}(t_0) &= \boldsymbol{z}_0 \\ \dot{\boldsymbol{y}}(t_0) &= \mathbf{0} \\ \ddot{\boldsymbol{y}}(t_0) &= \mathbf{0} \ . \end{aligned} \tag{4.112}$$

By introducing these equations into the equations of motion (4.97) of the rotating system one obtains a set of algebraic condition equations to be met:

$$\bar{n}^2 \begin{pmatrix} z_{0_1} \\ z_{0_2} \\ 0 \end{pmatrix} - k^2 \left\{ m_0 \, \frac{\boldsymbol{z}_0 - \boldsymbol{y}_0}{r_0^3} + m_1 \, \frac{\boldsymbol{z}_0 - \boldsymbol{y}_1}{r_1^3} \right\} = \mathbf{0} \ . \tag{4.113}$$

The structure of the above equations is recognized more easily, if they are written component-wise and if the representations (4.90, 4.91) are used for the coordinates of the two point masses m_0 and m_1:

$$\begin{aligned} &\bar{n}^2 \, z_{0_1} - k^2 \left\{ \frac{m_0}{r_0^3} \left[z_{0_1} + \frac{m_1}{m_0 + m_1} \, \bar{a} \right] + \frac{m_1}{r_1^3} \left[z_{0_1} - \frac{m_0}{m_0 + m_1} \, \bar{a} \right] \right\} &&= 0 \\ &\left\{ \bar{n}^2 - k^2 \left[\frac{m_0}{r_0^3} + \frac{m_1}{r_1^3} \right] \right\} z_{0_2} &&= 0 \\ &- k^2 \left[\frac{m_0}{r_0^3} + \frac{m_1}{r_1^3} \right] z_{0_3} &&= 0 \ . \end{aligned} \tag{4.114}$$

Stationary solutions only result, if all three eqns. (4.114) hold. As the coefficient of the third of eqns. (4.114) is always negative, stationary solutions are only possible for

$$z_{0_3} = 0 \ , \tag{4.115}$$

implying that stationary solutions have to lie in the (x, y)-plane, i.e., in the orbital plane of the two finite masses.

We make the distinction of two kinds of stationary solutions:

- For $z_{0_2} = 0$ candidate solutions lie on the straight line defined by the two point masses m_0 and m_1. Solutions of this kind are referred to as *straight line solutions* or *collinear point mass solutions.* The first of equations (4.114) provide the required condition equation. The straight line solutions were discovered by Euler.

- For $z_{0_2} \neq 0$ the point masses m_0, m_1 and the test particle have to form a (non-degenerated) triangle. One usually speaks of triangular solutions of the problème restreint. These solutions were discovered by Lagrange. In this case the coefficient of the term z_{0_2} in eqns. (4.114) must be zero:

$$\bar{n}^2 - k^2 \left\{ \frac{m_0}{r_0^3} + \frac{m_1}{r_1^3} \right\} = 0 \ . \tag{4.116}$$

  In addition, the first of the three equations (4.114) must hold. Let us now study the straight line and the triangular solutions separately.

Euler's Straight Line Solutions. Obviously, we have to assume that $z_{0_2} = 0$ and $z_{0_3} = 0$. Figure 4.17 shows that the test particle may assume three different positions w.r.t. the two point masses m_0 and m_1. It may either lie between the two bodies (position 1), to the right of the two bodies (position 2), or to the left of the two bodies (position 3). Independently of the particular position one may express r_1 in units of $\bar{a}$:

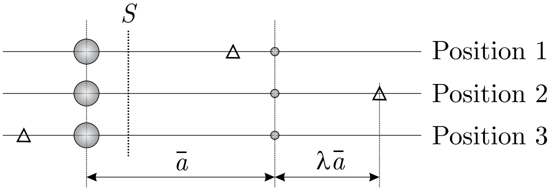


Fig. 4.17. Three positions of test particle w.r.t. the two point masses

$$r_1 \overset{\text{def}}{=} \lambda \, \bar{a} \ . \tag{4.117}$$

The dimension-free quantity λ may assume different values according to the position (relative to m_0 and m_1) considered:

$$\begin{array}{llll}
\text{Position 1:} & \lambda \in (0,1) \ , & r_0 = (1-\lambda)\,\bar{a} \ , & z_{0_1} = \left\{ \dfrac{m_0}{m_0+m_1} - \lambda \right\} \bar{a} \\
\text{Position 2:} & \lambda \in (0,\infty) \ , & r_0 = (1+\lambda)\,\bar{a} \ , & z_{0_1} = \left\{ \dfrac{m_0}{m_0+m_1} + \lambda \right\} \bar{a} \\
\text{Position 3:} & \lambda \in (1,\infty) \ , & r_0 = (\lambda-1)\,\bar{a} \ , & z_{0_1} = \left\{ \lambda - \dfrac{m_0}{m_0+m_1} \right\} \bar{a} \ .
\end{array} \tag{4.118}$$

Let us now assume that the considered body resides at position 1. If we replace the distances r_0, r_1 and z_{0_1} in the first of eqns. (4.114) according to eqn. (4.117) and (4.118) we obtain after division by $\bar{n}^2$ (and taking into account that $\bar{n}^2 \, \bar{a}^3 = k^2 \, (m_0 + m_1)$) the following condition equation for λ:

$$m_0 \; - \; (m_0 + m_1)\,\lambda \; - \; \frac{m_0}{(1-\lambda)^2} \; + \; \frac{m_1}{\lambda^2} = 0 \; . \qquad (4.119)$$

Let us define the function

$$K(\lambda) \overset{\text{def}}{=} m_0 \; - \; (m_0 + m_1)\,\lambda \; - \; \frac{m_0}{(1-\lambda)^2} \; + \; \frac{m_1}{\lambda^2} \qquad (4.120)$$

to simplify the subsequent discussion.

One easily verifies that $K(\lambda) \to +\infty$ for $\lambda \to 0$ and that $K(\lambda) \to -\infty$ for $\lambda \to 1$. This implies that the above condition equation has at least one root in the interval $I = (0, 1)$. As $K(\lambda)$ monotonically decreases in the same interval (take the derivative of $K(\lambda)$ w.r.t. λ) one may even conclude that there is exactly one root in the interval mentioned.

In an analogous way one may prove that there exists exactly one root for each of the condition equations related to the positions 2 and 3 of the considered test particle (see Figure 4.17).

In summary, we may state that there are three straight line solutions, one for each of the positions of the three bodies as shown in Figure 4.17. For $\frac{m_1}{m_0} \to 0$ the roots corresponding to the positions 1 and 2 approach $r_1 \to 0$ implying that the test particle becomes a satellite of mass m_1. Under the same assumption $r_1 \to 2\,\bar{a}$ for the position 3, and the test particle lies diametrally opposite to m_1 as seen from the point mass m_0. The semi-major axis of the orbit is (approximately) the same as that of m_1, but the orbital position is opposite to m_1 as seen from m_0.

Lagrange's Triangular Solutions. For (true) triangular solutions the first of eqns. (4.114) and eqn. (4.116) must hold. A simple rearrangement of terms in eqns. (4.114) leads to the result:

$$\left\{ \bar{n}^2 \; - \; k^2 \left[\frac{m_0}{r_0^3} + \frac{m_1}{r_1^3} \right] \right\} z_{0_1} \; - \; k^2 \, \frac{m_0\, m_1}{m_0 + m_1} \; \bar{a} \left\{ \frac{1}{r_0^3} - \frac{1}{r_1^3} \right\} = 0 \; . \quad (4.121)$$

According to the condition equation (4.116) the coefficient of z_{0_1} must be zero. This allows us to conclude immediately that

$$r_0 = r_1 \qquad (4.122)$$

must hold. If we use this result in eqn. (4.116) and take into account that $k^2\,(m_0 + m_1) = \bar{n}^2\,\bar{a}^3$ we obtain the final result

$$r_0 = r_1 = \bar{a} \; , \qquad (4.123)$$

which implies that the three celestial bodies of the problème restreint have to form an equilateral triangle in the orbital plane of the two bodies m_0 and m_1.

We were able to identify *five stationary solutions of the problème restreint*, namely three *straight line solutions* and two *triangular solutions.* Figure 4.18 illustrates the five solutions, where a mass ratio of $\frac{m_0}{m_1} = 3\,,\ m_0 + m_1 = 1\,,$ was used to draw the straight line solutions. The positions are labelled $L_i\,,\ i = 1,2,3,4,5$. They are also called libration points (implying that real celestial bodies would librate about these points). The letter L also reminds of Lagrange, who contributed significantly to the solution of the problème restreint. The stationary solutions of the poblème restreint are obviously particular "analytical" solutions (solutions in closed form) of the three body solution. Solutions in closed analytical form corresponding to the five solutions L_1 to L_5 may be found under more general conditions (all three masses different from zero, elliptic instead of circular orbits for m_0 and m_1). For more information we refer to [117], [49] (in German), or [31] (in English).

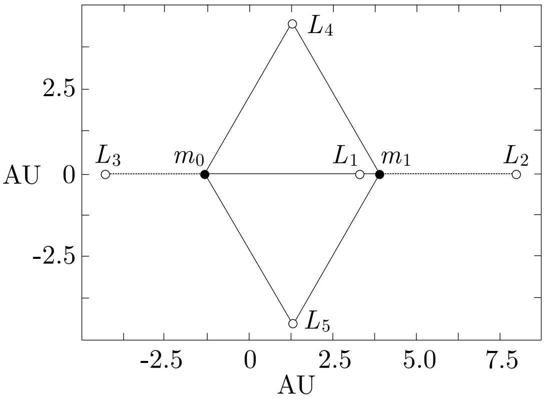


Fig. 4.18. The stationary solutions $L_i\,,\ i = 1,2,\ldots,5$, of the problème restreint

Stability of Stationary Solutions. Five stationary solutions of the problème restreint were found in the previous paragraph. Here we want to answer the question whether these solutions are *stable*, i.e., whether the orbit of a celestial body with slightly modified initial values will stay in the vicinity of the points L_i or not.

The tools to be used for this purpose will be fully developed in Chapter 5. Here we only need a very limited subset of the theory presented in that chapter.

Let

$$p \in \{z_{0_1}, z_{0_2}, z_{0_3}, \dot{z}_{0_1}, \dot{z}_{0_2}, \dot{z}_{0_3}\} \tag{4.124}$$

be one of the coordinates of the initial values of a particular solution of the equations of motion (4.97). Let us furthermore introduce the function

$$\boldsymbol{w}(t) \overset{\text{def}}{=} \frac{\partial \boldsymbol{y}}{\partial p}(t) \ . \tag{4.125}$$

To a first order, $\boldsymbol{w}(t)\,\Delta p$ tells what the consequence at time t of a change Δp in a particular initial value at time t_0 will be. Obviously the solution will remain in the neighborhood of the libration point considered, if $\boldsymbol{w}(t)$ shows only periodic variations. If this is true for all initial conditions (4.124) considered, the corresponding solution is said to be stable.

An ordinary differential equation system is obtained for the function $\boldsymbol{w}(t)$ by taking the partial derivative of the equations of motion (4.97) w.r.t. p (one of the initial values):

$$\ddot{\boldsymbol{w}} = \mathbf{A}_0\,\boldsymbol{w} \;+\; \mathbf{A}_1\,\dot{\boldsymbol{w}} \ . \tag{4.126}$$

The system (4.126) is a linear, homogeneous system of equations. It is referred to as the *system of variational equations* associated with the stationary solutions of the problème resteint. (A general derivation for the variational equations will be given in Chapter 5.1). The matrices $\mathbf{A}_i$, $i = 0, 1$ are obtained as:

$$\begin{aligned}\mathbf{A}_0 = \bar{n}^2 \begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,0 \end{pmatrix} \;&-\; k^2\,\frac{m_0}{r_0^3}\left\{\mathbf{E} \;-\; \frac{3}{r_0^2}\,(\boldsymbol{y}-\boldsymbol{y}_0)\otimes(\boldsymbol{y}-\boldsymbol{y}_0)^T\right\} \\ &-\; k^2\,\frac{m_1}{r_1^3}\left\{\mathbf{E} \;-\; \frac{3}{r_1^2}\,(\boldsymbol{y}-\boldsymbol{y}_1)\otimes(\boldsymbol{y}-\boldsymbol{y}_1)^T\right\} \ ,\end{aligned} \tag{4.127}$$

where $\otimes$ stands for the outer product of two vectors, and where

$$\mathbf{A}_1 = -\,2\,\bar{n} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ . \tag{4.128}$$

For the five stationary solutions to be considered, the radii r_0 and r_1 are constant, which in turn implies that the matrices $\mathbf{A}_i$, $i = 0, 1$, are time-independent, as well. For stationary solutions the system of variational equations is a linear, homogeneous system with constant coefficients – (almost) the simplest case of an ordinary differential equation system.

Stability of Triangular Solutions. From Figure 4.18 one obtains:

$$\boldsymbol{y}-\boldsymbol{y}_0 = \frac{\bar{a}}{2}\begin{pmatrix} +1 \\ \pm\sqrt{3} \\ 0 \end{pmatrix} \ , \tag{4.129}$$

as well as

$$\boldsymbol{y}-\boldsymbol{y}_1 = \frac{\bar{a}}{2}\begin{pmatrix} -1 \\ \pm\sqrt{3} \\ 0 \end{pmatrix} \ , \tag{4.130}$$

where the positive sign applies to the solution L_4, the negative sign to solution L_5 (see Figure 4.18).

Taking into account that $r_0 = r_1 = \bar{a}$ for the triangular solutions, matrix $\mathbf{A}_0$ reads as

$$\mathbf{A}_0 = \frac{3\,\bar{n}^2}{4} \begin{pmatrix} 1, & \pm\sqrt{3}\,\chi, & 0 \\ \pm\sqrt{3}\,\chi, & 3, & 0 \\ 0, & 0, & -\frac{4}{3} \end{pmatrix} , \tag{4.131}$$

where:

$$\chi = \frac{m_0 - m_1}{m_0 + m_1} . \tag{4.132}$$

The variational equations for triangular solutions thus read as

$$\ddot{\boldsymbol{w}} - \frac{3\,\bar{n}^2}{4} \begin{pmatrix} 1, & \pm\sqrt{3}\,\chi, & 0 \\ \pm\sqrt{3}\,\chi, & 3, & 0 \\ 0, & 0, & -\frac{4}{3} \end{pmatrix} \boldsymbol{w} + 2\,\bar{n} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \dot{\boldsymbol{w}} = \mathbf{0} . \tag{4.133}$$

Obviously the system is separated into one for the first two components and a scalar equation for the third equation

$$\ddot{w}_3 + \bar{n}^2\, w_3 = 0 . \tag{4.134}$$

This is the equation of the one-dimensional harmonic oscillator which is solved by:

$$w_3(t) = w_{0_3}\, \cos(\bar{n}t + \alpha) , \tag{4.135}$$

where w_{0_3} is the initial value of $w_3(t)$ at $t = 0$. Obviously the general solution is periodic in nature. The triangular solutions at L_4 and L_5 are therefore stable w.r.t. small changes in the third component of the initial position and the velocity vectors.

The variational equations for the components $w_1(t)$ und $w_2(t)$ are (first two of eqns. (4.133)):

$$\begin{pmatrix} \ddot{w}_1 \\ \ddot{w}_2 \end{pmatrix} + 2\,\bar{n} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \dot{w}_1 \\ \dot{w}_2 \end{pmatrix} - \frac{3\,\bar{n}^2}{4} \begin{pmatrix} 1, & \pm\sqrt{3}\,\chi \\ \pm\sqrt{3}\,\chi, & 3 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \mathbf{0} . \tag{4.136}$$

The system is solved by

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = e^{\lambda\,(t-t_0)} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} , \tag{4.137}$$

where λ is an arbitrary complex number and c_1 and c_2 are (arbitrary) constants. Should λ be purely imaginary only periodic solutions are obtained. In all other cases there will be components showing an exponential growth. Obviously the solutions would be stable in the former, unstable in the latter case.

Using eqn. (4.137) in the variational equations (4.136) one obtains the algebraic condition equations

$$\begin{pmatrix} \lambda^2 - \frac{3}{4}\,\bar{n}^2 & , -\bar{n}\left(2\,\lambda \pm \frac{3\,\sqrt{3}}{4}\,\bar{n}\,\chi\right) \\ +\bar{n}\left(2\,\lambda \mp \frac{3\,\sqrt{3}}{4}\,\bar{n}\,\chi\right) , & \lambda^2 - \frac{9}{4}\,\bar{n}^2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \mathbf{0} \ . \tag{4.138}$$

Non-trivial solutions are obtained only if the determinant of the above 2 × 2-matrix is identically zero. This requirement leads to the following biquadratic equation in λ

$$\lambda^4 + \bar{n}^2\,\lambda^2 + \tfrac{27}{16}\,\bar{n}^4\left(1 - \chi^2\right) = 0 \ , \tag{4.139}$$

which is solved by:

$$\lambda^2 = -\frac{\bar{n}^2}{2}\left\{1 \pm \sqrt{1 - \tfrac{27}{4}\left(1 - \chi^2\right)}\right\} \ . \tag{4.140}$$

The four solutions obviously are purely imaginary if

$$0 \le 1 - \tfrac{27}{4}\left(1 - \chi^2\right) < 1 \ . \tag{4.141}$$

Assuming that $m_0 \ge m_1$ this implies $0 \le \chi \le 1$. This in turn means that stable solutions are obtained if the mass ratio meets the condition

$$\chi^2 \ge \tfrac{23}{27} \tag{4.142}$$

or

$$\chi = \frac{m_0 - m_1}{m_0 + m_1} \ge 0.9229582 \ . \tag{4.143}$$

Triangular solutions are obviously stable if $m_0 \gg m_1$. For the three-body problem Sun-Jupiter-Minor planet the triangular solutions are obviously stable, because we have

$$\chi_{\odot\jupiter} = \frac{1 - \frac{1}{1047.35}}{1 + \frac{1}{1047.35}} = 0.9980922 > 0.9929582 \ . \tag{4.144}$$

In Chapter II- 4 we will see that in the solar system the libration points L_4 and L_5 (of the three-body problem Sun-Jupiter-asteroid) are populated by the Trojans (and Greek) family of asteroids. The existence of these families of minor planets underlines the importance of the problème restreint in practice.

From the above equations one easily derives the following basic frequencies of the system Sun-Jupiter-Minor planet:

$$\begin{aligned} \lambda_{1,2} &= \nu_{1,2}\ i = \pm\,0.9967575\,\bar{n}\,i \\ \lambda_{3,4} &= \nu_{3,4}\ i = \pm\,0.0804641\,\bar{n}\,i \ , \end{aligned} \tag{4.145}$$

where i is the imaginary unit. In essence we have one period which is comparable to the revolution period of the two masses m_0 and m_1, and one which is about twelve times longer. The latter period is observed as a slow librational motion of the Trojan (Greek) family of minor planets about the libration points L_4 and L_5.

Instability of Straight Line Solutions. We follow the procedure established for the triangular solutions to address the issue of stability of the straight line solutions. In a first step the variational equations (4.126) have to be formulated for the special case considered. We observe that:

$$\boldsymbol{y}-\boldsymbol{y}_0=\begin{pmatrix} r_0\\ 0\\ 0\end{pmatrix} \quad \text{and} \quad \boldsymbol{y}-\boldsymbol{y}_1=\begin{pmatrix} r_1\\ 0\\ 0\end{pmatrix} \ . \tag{4.146}$$

Let us select the libration point L_2 as an example. Matrix $\mathbf{A}_0$ assumes the form:

$$\mathbf{A}_0=\bar{n}^2\begin{pmatrix} 1+2\,\zeta^2\ , & 0 & , \ 0\\ 0 & ,\ 1-\zeta^2\ , & 0\\ 0 & ,\ 0 & ,\ -\zeta^2\end{pmatrix} \ , \tag{4.147}$$

where

$$\zeta^2=\frac{m_0}{m_0+m_1}\,\frac{1}{(1+\lambda)^3}+\frac{m_1}{m_0+m_1}\,\frac{1}{\lambda^3}>0 \ . \tag{4.148}$$

With these relations the variational equations relative to the stationary solution at L_2 are obtained as

$$\ddot{\boldsymbol{w}}\,-\,\bar{n}^2\begin{pmatrix} 1+2\,\zeta^2, & 0, & 0\\ 0, & 1-\zeta^2 & 0\\ 0, & 0, & -\zeta^2\end{pmatrix}\boldsymbol{w}\,+\,2\,\bar{n}\begin{pmatrix} 0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\dot{\boldsymbol{w}}=\mathbf{0} \ . \tag{4.149}$$

Obviously, it is again possible (as in the case of the triangular solutions) to split up the system into an equation for the third component (normal to the orbital plane of the bodies m_0 and m_1) and a system of equations for the first two components of the vector $\boldsymbol{w}$. According to the above definition the auxiliary scalar quantity ζ is real, which is why we may conclude that $\zeta^2>0$. This in turn implies that the third component $w_3(t)$ obeys the equation

$$\ddot{w}_3\,+\,\zeta^2\,\bar{n}^2\,w_3=0 \tag{4.150}$$

of the harmonic oscillator, the solutions of which are periodic functions. The same result (with a slightly different period) was obtained for the triangular solutions. The considered straight-line solution thus is stable w.r.t. small changes in the third components of the initial position and velocity vectors.

The first two components obey the following system of equations:

$$\begin{pmatrix}\ddot{w}_1\\ \ddot{w}_2\end{pmatrix}+2\,\bar{n}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\dot{w}_1\\ \dot{w}_2\end{pmatrix}-\bar{n}^2\begin{pmatrix}1+2\,\zeta^2, & 0\\ 0, & 1-\zeta^2\end{pmatrix}\begin{pmatrix}w_1\\ w_2\end{pmatrix}=\mathbf{0}\ , \tag{4.151}$$

which is solved by the same type (4.137) of solution as in the case of the triangular solutions. Introducing formula (4.137) into the variational equations (4.151) for the libration point L_2 yields the following condition equations for the coefficient λ:

$$\begin{pmatrix} \lambda^2 - (1 + 2\,\zeta^2)\,\bar{n}^2 \;, & -\,2\,\bar{n}\,\lambda \\ 2\,\bar{n}\,\lambda & ,\; \lambda^2 - (1 - \zeta^2)\,\bar{n}^2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \mathbf{0}\ . \qquad (4.152)$$

Non trivial solutions only exist if the determinant of this matrix is zero. One easily verifies that the above 2×2 matrix gives rise to a biquadratic expression in λ for the determinant. Also one may show in a straight forward way that it is *not* possible in this case to find purely imaginary solutions. Therefore, there are *no* purely periodic solutions of the variational equations (4.151) for the straight-line solution considered. This in turn implies that the solution of the equations of motion (4.97) with initial conditions slightly (infinitesimally) different from those of the stationary solution at L_2 will eventually depart exponentially from L_2. Obviously, the stationary solution at L_2 is not stable.

One easily verifies that the solutions corresponding to the other two libration points L_1 and L_3 are *not* stable either. Before the advent of the space flight era one could therefore safely state that only the libration points L_4 and L_5 are of *practical importance* (family of asteroids associated with them in the three-body problem Sun-Jupiter-asteroid). The libration points L_1 and L_2 of the three-body-problem *Sun-Earth-spacecraft* are of considerable importance today, e.g., for Sun-observing space missions: when brought to either of the two libration points, the spacecraft will remain in the vicinity of the Earth (observe the mass ratio $m_{\odot} : m_{♁}$) and it will co-rotate with the Earth around the Sun (the geocentric and heliocentric revolution periods are the same). The stability of the solution, as considered in this section, is not a central issue for spaceflight applications, because orbital manoeuvres are anyway necessary (e.g., due to non-gravitational forces) to keep the spacecrafts in place.

Periodic Solutions in the Problème Restreint. We were looking for stationary solutions in the rotating system in the previous paragraphs. We might as well study periodic solutions (and their stability) in the same coordinate system. This would immediately lead to the question whether there are such solutions and (in the affirmative case) how they can be established in practice. A thorough discussion requires the distinction of many special cases, e.g., periodic orbits about m_0, m_1, the libration points L_4 and L_5, periodic orbits around m_0 *and* m_1, etc. We refer to [117] for a broad discussion of periodic solutions.

H. Poincaré was very much interested in the periodic solutions of the problème restreint. He hoped to obtain clues concerning the stability of the solar system with their help. He had the idea to associate a periodic orbit with each real orbit (close to the latter) and to study stability properties with the help of the periodic orbit (which, by construction, would be rather simple to describe) and the variational equations associated with it. This concept eventually failed because he found that it was not even possible to find a periodic solution for each revolution period, or, what is equivalent, for each mean motion n of a test particle.

Let us conclude the chapter by constructing a few periodic orbits around m_0 for a planetary three-body problem, i.e., for $m_0 \gg m_1$. (The actual values for the problème resteint Sun-Jupiter-asteroid will be used below). We focus on orbits with a period $\tilde{P} \overset{\text{def}}{=} \frac{2\pi}{n-\bar{n}}$ in the rotating system. Obviously n must be the mean motion of the asteroid as observed in the inertial system, $\bar{n}$ is the orbital period of Jupiter around the Sun.

Periodic orbits of this kind may be generated numerically by defining a parameter estimation problem. It is (in general) safe to assume, that a circular orbit with the mean motion n is a fair approximation to the true solution (it approximates the true solution for $m_1 \to 0$). With this assumption we may define approximate values for the initial conditions of a truly periodic orbit in the rotating system: The initial epoch $t_0 = 0$ is chosen to correspond to a crossing of the first coordinate axis in the rotating system. As a circular orbit was assumed, the initial conditions read as:

$$\boldsymbol{y}(0) = \begin{pmatrix} y^*_{0_1} \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \dot{\boldsymbol{y}}(0) = \begin{pmatrix} 0 \\ \dot{y}^*_{0_2} \\ 0 \end{pmatrix} , \tag{4.153}$$

where

$$y^*_{0_1} \overset{\text{def}}{=} \left(\frac{k^2\, m_0}{n^2} \right)^{\frac{1}{3}} \quad \text{and} \quad \dot{y}^*_{0_2} = y^*_{0_1}\,(n - \bar{n}) \ . \tag{4.154}$$

A truly periodic orbit in the rotating system with the period $\tilde{P} = \frac{2\pi}{n-\bar{n}}$ must intersect the first coordinate axis after the time period $\Delta t = \frac{\tilde{P}}{2}$ perpendicularly: For reasons of symmetry, the orbit in the time interval $[\frac{1}{2}\tilde{P}, \tilde{P}]$ will then be a mirror (with the first coordinate axis as a mirror) of the orbit in the time interval $[0, \frac{1}{2}\tilde{P}]$. A truly periodic solution $\boldsymbol{y}_p(t)$ therefore must meet the following conditions:

$$y_{p,02}\left(\frac{\tilde{P}}{2}\right) \overset{\text{def}}{=} 0 \quad \text{and} \quad \dot{y}_{p,01}\left(\frac{\tilde{P}}{2}\right) \overset{\text{def}}{=} 0 \ . \tag{4.155}$$

With the approximate initial conditions defined above these conditions will be met only approximately. How do we establish it to generate an orbit meeting the condition equations (4.154) precisely? The answer is simple: we represent the unknown, truly periodic orbit by

$$\boldsymbol{y}_p(t) \overset{\text{def}}{=} \boldsymbol{y}(t) \,+\, \boldsymbol{w}_1\, \Delta y^*_{0_1} \,+\, \boldsymbol{w}_2\, \Delta \dot{y}^*_{0_2} \ , \tag{4.156}$$

where $\boldsymbol{w}_i$, $i = 1, 2$, are the solutions of the variational equations (4.126) w.r.t. the initial values $y^*_{0_1}$ and $\dot{y}^*_{0_2}$. These two initial values must now be varied in such a way that the orbit $\boldsymbol{y}_p(t)$ meets the two condition equations (4.155). Introducing the orbit representation (4.156) into the condition equations (4.155) we obtain the following system of two linear equations for the determination of improved initial conditions:

$$\begin{aligned} y_{p,02}\left(\frac{\tilde{P}}{2}\right) &\overset{\text{def}}{=} y_2\left(\frac{\tilde{P}}{2}\right) + w_{1,2}\left(\frac{\tilde{P}}{2}\right)\Delta y_{0_1} + w_{2,2}\left(\frac{\tilde{P}}{2}\right)\Delta \dot{y}_{0_2} = 0 \\ \dot{y}_{p,01}\left(\frac{\tilde{P}}{2}\right) &\overset{\text{def}}{=} \dot{y}_1\left(\frac{\tilde{P}}{2}\right) + \dot{w}_{1,1}\left(\frac{\tilde{P}}{2}\right)\Delta y_{0_1} + \dot{w}_{2,1}\left(\frac{\tilde{P}}{2}\right)\Delta \dot{y}_{0_2} = 0\ . \end{aligned} \tag{4.157}$$

The solution of the two linear equations (4.157) are used to define new initial values $y_{0_1} \overset{\text{def}}{=} y^*_{0_1} + \Delta y_{0_1}$ and $\dot{y}_{0_2} \overset{\text{def}}{=} \dot{y}^*_{0_2} + \Delta \dot{y}_{0_2}$ which will result in an orbit meeting the conditions (4.155) much better than the original orbit represented by the initial value (4.153). Should there still be unacceptable discrepancies, the parameter estimation procedure outlined above may be repeated with the improved initial values as a priori approximations.

The sketched parameter estimation process will in general have a unique solution – two scalar condition equations are necessary and sufficient to determine two free parameters, which may be adjusted.

The resulting orbits usually are close to circular. Exceptions to this rule occur if the revolution periods of the test particle and the two bodies m_0 and m_1 are commensurable. Two orbital periods $\bar{P}$ and P are said to be commensurable if their periods (in the inertial system) may be expressed by

$$\frac{\bar{P}}{P} = \frac{k_1}{k_2}\ , \tag{4.158}$$

where k_1 and k_2 are integers. Observe, that commensurable periodic orbits in the rotating system are also periodic in the inertial system (what is not true in the general case).

The parameter estimation process (4.157) will fail, if the periods of the minor planet and Jupiter (as measured in the inertial system) meet the additional condition

$$P = \frac{k-1}{k}\,\bar{P}\ , \quad k = 2, 3, \dots\ . \tag{4.159}$$

Figure 4.19 (left) shows two periodic orbits near the (2 : 1)-commensurability, Figure 4.19 (right) two periodic orbits near the (3 : 2)-commensurability. The differences in the mean motions of the two orbits are very small in both cases, the resulting solutions differ dramatically, however. Moreover, the solutions are highly eccentric. In the example of Figure 4.19 (left) the perihelia of the two orbits differ by 90° in the rotating system (the close encounters of Jupiter and the minor planet occur at perihelia and at aphelia passing times of the minor planet, respectively). Note that there is a group of minor planets, the so-called Hilda group. The solid-line solution of Figure 4.19 (right) corresponds to the Hilda-type objects (see Chapter II- 4, section II- 4.3.4 for more information). The closer one tries to approach the exact commensurabilities, the more eccentric the corresponding periodic orbits become.

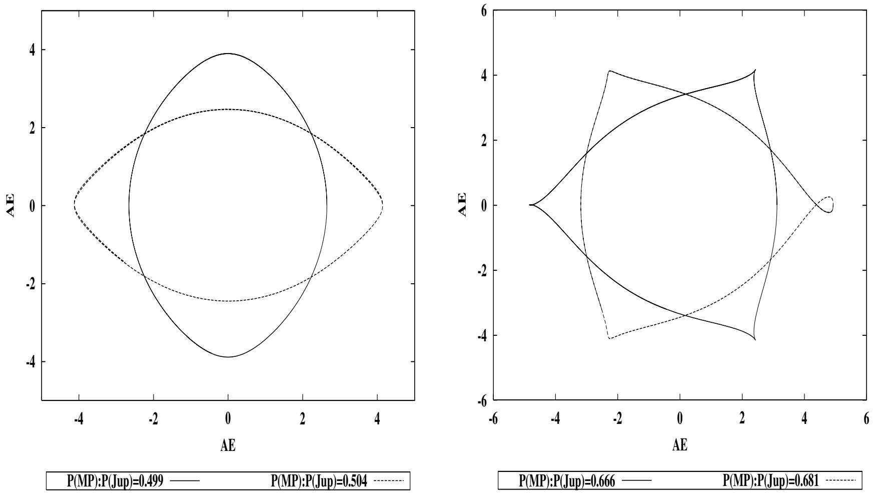


Fig. 4.19. Two periodic orbits near to the (2 : 1)- and to the (3 : 2)-commensurability of three-body problem Sun-Jupiter-minor planet

This behavior also implies that already in the case of the problème restreint there are orbits which are chaotic, i.e., even if the initial conditions of two orbits differ only by infinitesimally small amounts, the resulting solutions will depart from each other exponentially. Poincaré's investigations of the periodic solutions in the problème restreint were of fundamental importance. They eventually led to the developments of the theory of dynamical systems and to the detection of the so-called deterministic chaos, which will be discussed in greater detail in Chapter II- 4.

The above figures were generated with the program PLASYS (documented in Chapter II- 10).

5. Variational Equations

5.1 Motivation and Overview

When discussing the stability of stationary solutions in the problème restreint in section 4.5.2 we derived differential equations for the partial derivatives of these solutions w.r.t. the components of the initial state vector and called them *variational equations.*

Variational equations may be attributed to all trajectories (particular solutions) solving the different types of equations of motion given in Chapter 3. As a matter of fact, variational equations exist for all initial (and boundary) value problems – inside and outside the area of Celestial Mechanics. In this chapter we review the essential properties of the variational equations encountered in Celestial Mechanics. These equations are in particular indispensable for

- stability studies of particular solutions of differential equations (see sections 4.5.2, II- 4.3.4 for applications),
- parameter estimation theory, in particular orbit determination (see Chapter 8), and
- the theory of error propagation in numerical integration (see Chapter 7).

Variational equations are of central importance for the theory of dynamical systems, in particular in Celestial Mechanics.

In section 5.2 the problem is studied from a rather general point of view. We depart from an initial value problem associated with a differential equation system of dimension d and order $n \geq 1$ and derive the general form of the variational equations. As the two-body problem could be solved in closed form, it must also be possible to solve the corresponding variational equations "analytically". This task is accomplished in section 5.3, where extensive use is made of the formulas of the two-body problem in Chapter 4. The variational equations accompanying the perturbed trajectory of one celestial body are studied in section 5.4, those associated with the general N-body problem in section 5.5. Efficient solution methods for the variational equations occurring in Celestial Mechanics are based on the (analytically known) solutions of the corresponding equations of the two-body problem. Methods of this type are outlined in section 5.6. In the concluding section 5.7 the variational equations are used to study the impact of small errors introduced at a particular time t_k on the trajectory for $t \geq t_k$. The so-called fundamental law of error propagation is derived in this section. Unnecessary to say that this law is of vital importance when studying the propagation of numerical integration errors in Chapter 7.

5.2 Primary and Variational Equations

Primary Equations. The three types of equations of motion derived in Chapter 3, namely the equations of the N-body problem, the three-body problem Earth-Moon-Sun, and of the motion of artificial Earth satellites, are non-linear systems of the second order. In Chapter 6 we will see that it is equally well possible to transform the equations of motion into a system of first order with the osculating orbital elements as dependent arguments. Therefore we derive the variational equations associated with the following general initial value problem:

$$\begin{aligned} \boldsymbol{y}^{(n)} &= \boldsymbol{f}(t; \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}, \tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_m) \\ \boldsymbol{y}^{(i)}(t_0) &= \boldsymbol{y}_0^{(i)} \;, \quad i = 0, 1, \ldots, n-1 \;, \end{aligned} \tag{5.1}$$

where:

$n \geq 1$ is the order of the differential equation system, d its dimension;

$\boldsymbol{y}(t)$, a column array with d elements, is the solution vector of the system,

$\dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(i)}$, represent the first, second, ... ith derivative of the solution vector,

$\boldsymbol{y}_0^{(i)}$, $i = 0, 1, \ldots, n-1$, is the initial state vector of the system at the initial epoch t_0 (the initial state vector may be understood as the set of the nd components of the solution vector and its first $n-1$ derivatives at $t = t_0$) – these components are also referred to as "initial values";

$\boldsymbol{f}(\ldots)$, a column vector with d elements, is the right-hand side of the differential equation system, and

$\tilde{p}_j$, $j = 1, 2, \ldots, m$, are the so-called dynamical parameters of the system.

In the context of variational equations the initial value problem (5.1) is referred to as the system of *primary equations.*

The dynamical parameters are included explicitly in eqns. (5.1), because one may wish to study the impact of (slightly) changing one or more of these parameters on the particular solution of the primary equations. In this case the partial derivatives of the particular solution defined by eqns. (5.1) w.r.t.

these parameters are required in addition to the partials w.r.t. the components of the initial state vector. The masses of the planets in eqns. (3.18) or the coefficients C_{ik}, S_{ik} of the Earth's gravitational potential in the equations of motion (3.143, 3.150) are examples of dynamical parameters.

It is possible to use more general parameters describing the initial state of the system than the components of the vectors $\boldsymbol{y}_0^{(i)}$, $i = 0, 1, \ldots, n-1$. This is, as a matter of fact, frequently done: In Chapter 8 we will, e.g., use the osculating elements at t_0 to describe the initial state vector. As this more general treatment does not alter the structure of the problem, we stick to the formally simpler way of parametrization and consider the components of the above mentioned vectors as parameters.

It is even possible to replace the initial value problem (5.1) by a boundary value problem. This may, e.g., make sense when determining orbits, if the problem is formulated as a local boundary value problem (see Chapter 8). As it is a straightforward matter to generalize the tools developed subsequently to boundary value problems of this kind, we do "only" address variational equations associated with primary equations of type (5.1) from here onwards.

Notation. Let us introduce the following notation for the $nd{+}m$ parameters of the above primary equations (5.1)

$$\{p_1, p_2, \ldots, p_{nd+m}\} \overset{\text{def}}{=} \left\{ \boldsymbol{y}_0^T, \dot{\boldsymbol{y}}_0^T, \ddot{\boldsymbol{y}}_0^T, \ldots, \left(\boldsymbol{y}_0^{(n-1)}\right)^T, \tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_m \right\} \ . \quad (5.2)$$

Note that the first nd parameters p_i define the initial values, whereas the last m parameters characterize the differential equation system. With this notation the particular solution defined by eqns. (5.1) may also be written as $\boldsymbol{y}(t; p_1, p_2, \ldots, p_{nd+m})$, if one wishes to underline the dependency of the solution vector on the parameters (5.2).

Let us now assume that

$$p \in \{p_1, p_2, \ldots, p_{nd+m}\} \quad (5.3)$$

is anyone of the parameters defining the initial value problem (5.1). The partial derivative of the solution vector w.r.t. this parameter p is then designated by

$$\boldsymbol{z}(t) \overset{\text{def}}{=} \left(\frac{\partial \boldsymbol{y}}{\partial p} \right)(t) \ , \quad (5.4)$$

where $\boldsymbol{y}(t) = \boldsymbol{y}(t; p_1, p_2, \ldots, p_{nd+m})$ is the solution vector of the initial value problem (5.1).

Sometimes it is important to refer to a particular parameter p_l in the list (5.2). In this case we will designate the corresponding partial derivative by:

$$\boldsymbol{z}_l(t) \overset{\text{def}}{=} \left(\frac{\partial \boldsymbol{y}}{\partial p_l} \right)(t) \ , \quad l \in \{1, 2, \ldots, nd+m\} \ . \quad (5.5)$$

$\boldsymbol{z}(t)$ and $\boldsymbol{z}_l(t)$ are functions of time. Their first, second, ... derivatives w.r.t. time are denoted by $\boldsymbol{z}^{(i)}$, $\boldsymbol{z}_l^{(i)}$, $i = (0), 1, 2, \ldots, n$, (in analogy to the notation used for vector $\boldsymbol{y}(t)$).

Variational Equations. The function $\boldsymbol{z}(t)$ solves the initial value problem, which is obtained by taking the partial derivative w.r.t. the parameter p of all equations, the differential equations and the defining equations for the initial state vector, in the original initial value problem (5.1).

Making extensive use of the chain rule of elementary calculus (by taking first the partial derivatives of the differential equation in eqns. (5.1) w.r.t. the components of the solution vector and its first $n-1$ derivatives at time t, then the partial derivatives of these components w.r.t. the parameter p), one easily verifies that $\boldsymbol{z}(t)$ is a particular solution of the linear, in general inhomogeneous differential equation system

$$\boldsymbol{z}^{(n)}(t) = \sum_{i=0}^{n-1} \mathbf{A}_i(t)\, \boldsymbol{z}^{(i)}(t) \; + \; \boldsymbol{f}_p(t) \ , \tag{5.6}$$

where the components of the square matrices of dimension $d \times d$

$$A_{i,jk}(t) \overset{\text{def}}{=} \left(\frac{\partial f_j}{\partial y_k^{(i)}} \right)(t) \ ; \quad i = 0, 1, \ldots, n-1, \quad j, k = 1, 2, \ldots, d \tag{5.7}$$

are the elements of the Jacobian of vector $\boldsymbol{f}(\ldots)$ w.r.t. the components of the solution vector and its first $n-1$ derivatives, respectively. The column matrix $\boldsymbol{f}_p(t)$ is the explicit derivative, i.e., the partial derivative w.r.t. the parameter p considered, of vector $\boldsymbol{f}(\ldots)$ (disregarding the fact that the vectors $\boldsymbol{y}^{(i)}$, $i = 0, 1, \ldots, n-1$, also depend on p, see right-hand side of the differential equation in eqns. (5.1)):

$$\boldsymbol{f}_p \overset{\text{def}}{=} \begin{cases} \qquad \mathbf{0} \qquad , & p \in \{p_1, p_2, \cdots, p_{nd}\} \\ \left(\dfrac{\partial \boldsymbol{f}}{\partial p} \right)(t) \ , & p \in \{p_{nd+1}, p_{nd+2}, \cdots, p_{nd+m}\} \ . \end{cases} \tag{5.8}$$

This explicit derivative is the zero vector $\mathbf{0}$, if the parameter p considered is one of the initial values.

The differential equation system (5.6) is called the *system of variational equations* for parameter p of the original differential equation system (5.1) (of the primary equations).

The initial values associated with the variational equations (5.6) for a parameter p are obtained by taking the partial derivatives of the corresponding initial values of the primary equations (5.1) w.r.t. parameter p, i.e., by

$$\boldsymbol{z}^{(i)}(t_0) = \frac{\partial \boldsymbol{y}_0^{(i)}}{\partial p} \ , \quad i = 0, 1, \ldots, n-1 \ . \tag{5.9}$$

The right-hand sides of eqns. (5.9) must all be identically zero, if parameter p is one of the dynamical parameters (see parameter definition (5.2)):

$$\boldsymbol{z}^{(i)}(t_0) = \mathbf{0}, \quad i = 0, 1, \ldots, n-1, \quad p \in \{p_{nd+1}, p_{nd+2}, \ldots, p_{nd+m}\} \ . \tag{5.10}$$

If, on the contrary, parameter $p \overset{\text{def}}{=} y_{0k}^{(l)}$ corresponds to component k of an initial value $\boldsymbol{y}_0^{(l)}$, we obtain

$$\boldsymbol{z}^{(i)}(t_0) = \begin{cases} \mathbf{0} \ , & i \neq l \\ \boldsymbol{e}_k \ , & i = l \end{cases} \ , \quad i = 0, 1, \ldots, n-1 \ . \tag{5.11}$$

The array $\boldsymbol{e}_k$ therefore must be interpreted as the column array formed by column k of the identity matrix $\mathbf{E}$ with $d \times d$ elements.

Elementary Properties of the Variational Equations. The properties of the variational equations (5.6) and their solutions accompanying the primary problem (5.1) may be summarized as follows:

1. The system (5.6) of variational equations is a linear differential equation system.
2. One solution vector $\boldsymbol{z}(t)$ is associated with each of the $nd+m$ parameters p_l , $l = 1, 2, \ldots, nd+m$, (5.2).
3. The partial derivatives $\boldsymbol{z}_l(t)$, $l = 1, 2, \ldots, nd$, w.r.t. the initial values are particular solutions of one and the same *homogeneous* differential equation system

   $$\boldsymbol{z}_l^{(n)}(t) = \sum_{i=0}^{n-1} \mathbf{A}_i(t)\, \boldsymbol{z}_l^{(i)} \ , \quad l = 1, 2, \ldots, nd \ . \tag{5.12}$$

   The nd solutions obey nd different sets of initial conditions (5.11).
4. The nd solutions $\boldsymbol{z}_l(t)$ associated with the initial conditions define a *complete system of solutions of the homogeneous equations (5.12).*
5. Each linear combination of the nd solutions of the homogeneous system

   $$\boldsymbol{z}(t) = \sum_{l=1}^{nd} \alpha_l\, \boldsymbol{z}_l(t) \tag{5.13}$$

   (with constant coefficients α_l) is a solution of the homogeneous system (5.12), as well. Function (5.13) may be considered as the general solution of the homogeneous system (5.12).
6. The attribute “complete” of the system of nd solutions $\boldsymbol{z}_l(t)$ reflects the fact that *any* solution of the homogeneous system (5.12) may be expressed as a linear combination of type (5.13).

The homogeneous equations (5.12) occur in each of the problems mentioned initially: when determining an orbit in a known force field, we have to determine the initial values (or functions thereof), in stability theory we are (usually) interested in the sensitivity of a particular trajectory w.r.t. small changes in the initial values; when studying error propagation in numerical integration we have to study the impact of small changes in the initial values – at time t_0 and at intermediary epochs t_i.

Solution of the Inhomogeneous Variational Equation by Quadrature. The inhomogeneous variational equations (5.6) only have to be solved, if the force field contains dynamical parameters p_j , $j = nd{+}1, nd{+}2, \ldots, nd{+}m$. In satellite-geodetic problems many of these parameters may occur in the same parameter estimation problem. When determining the Earth's gravity field parameters C_{ik}, S_{ik} up to degree and order 300, about 90000 dynamical parameters have to be determined. It may, therefore, be of crucial importance to have efficient solution methods at hand for such purposes.

In this paragraph it is shown how the inhomogeneous equations (5.6) may be solved by quadrature. This transformation is important because the solution methods for solving definite integrals are much more powerful than the methods for solving an ordinary differential equation system.

The partial derivatives w.r.t. the dynamical parameters p_j , $j = nd+1, nd+2, \ldots, nd{+}m$, are solutions of the *inhomogeneous* differential equation system (5.6), where the explicit derivative has to be taken w.r.t. parameter p_j. The initial values at time t_0, according to eqns. (5.10), are homogeneous (all zero).

The solution vector $\boldsymbol{z}_j(t)$, $j \in \{nd+1, nd+2, \ldots, nd+m\}$ of the inhomogeneous system (5.6) (and its first $n-1$ derivatives) are, e.g., obtained by the method of *variation of constants* as a linear combination of the nd homogeneous solutions

$$\boldsymbol{z}_j^{(i)}(t) \overset{\text{def}}{=} \sum_{l=1}^{nd} \alpha_l(t)\, \boldsymbol{z}_l^{(i)}(t) \ , \quad i = 0, 1, \ldots, n-1 \ , \tag{5.14}$$

where the coefficients $\alpha_l(t)$ are functions of time t (to be determined below). The functions $\boldsymbol{z}_l(t)$ on the right-hand side of eqns. (5.14) are solutions of the homogeneous system, they form the complete system of homogeneous solutions of the equation (5.12).

Observe that eqns. (5.14) are rather special: If you would take the ith time derivative of the function $\boldsymbol{z}_j(t)$, you would obtain much more general formulas than eqns. (5.14). We will make use of eqns. (5.14) to set up linear differential equations for the parameters α_l. In order to do that we write the above equation in a very compact matrix form by introducing the one-dimensional array

$$\boldsymbol{\alpha}^T(t) \overset{\text{def}}{=} \big(\alpha_1(t), \alpha_2(t), \ldots, \alpha_{nd}(t)\big) \ , \tag{5.15}$$

with nd elements, which allows it to write eqns. (5.14) in the form

$$\boldsymbol{z}_j^{(i)}(t) = \mathbf{Z}^{(i)}(t)\,\boldsymbol{\alpha}(t) \overset{\text{def}}{=} \mathbf{Z}^{(i)}\,\boldsymbol{\alpha}\ , \quad i = 0, 1, \ldots, n-1\ , \tag{5.16}$$

where $\mathbf{Z}(t)$ is the rectangular matrix with nd columns and d rows, in which column l contains the elements of the solution $\boldsymbol{z}_l(t)$ with index l of the homogeneous system.

Equations (5.16) and the inhomogeneous equation (5.6) allow it to list nd scalar condition equations for the nd components of the array $\boldsymbol{\alpha}(t)$:

$$\begin{array}{llcll}
\dfrac{d}{dt}\,(\mathbf{Z}\,\boldsymbol{\alpha}) & = \dot{\mathbf{Z}}\,\boldsymbol{\alpha} & \to & \mathbf{Z}\,\dot{\boldsymbol{\alpha}} & = \mathbf{0} \\
\dfrac{d^2}{dt^2}\,(\mathbf{Z}\,\boldsymbol{\alpha}) & = \ddot{\mathbf{Z}}\,\boldsymbol{\alpha} & \to & \dot{\mathbf{Z}}\,\dot{\boldsymbol{\alpha}} & = \mathbf{0} \\
\ldots & = \ldots & \to & \ldots & = \ldots \\
\ldots & = \ldots & \to & \ldots & = \ldots \\
\dfrac{d^{(n-1)}}{dt^{(n-1)}}\,(\mathbf{Z}\,\boldsymbol{\alpha}) & = \mathbf{Z}^{(n-1)}\,\boldsymbol{\alpha} & \to & \mathbf{Z}^{(n-2)}\ \dot{\boldsymbol{\alpha}} & = \mathbf{0} \\
\dfrac{d^{(n)}}{dt^{(n)}}\,(\mathbf{Z}\,\boldsymbol{\alpha}) & = \mathbf{Z}^{(n)}\,\boldsymbol{\alpha} + \mathbf{Z}^{(n-1)}\,\dot{\boldsymbol{\alpha}} & & & \\
 & = \sum\limits_{i=0}^{n-1} \mathbf{Z}^{(i)}\,\boldsymbol{\alpha} + \boldsymbol{f}_{p_j} & \to & \mathbf{Z}^{(n-1)}\,\dot{\boldsymbol{\alpha}} & = \boldsymbol{f}_{p_j}\ ,
\end{array} \tag{5.17}$$

where $\boldsymbol{f}_{p_j} = \left(\frac{\partial \boldsymbol{f}}{p_j}\right)(t)\,, \quad j \in \{nd+1, nd+2, \ldots, nd+m\}$.

For the sake of clarity we include the intermediary steps leading to the latter condition equation. When forming this equation we already know (assumed) that

$$\frac{d^{(n-1)}}{dt^{(n-1)}}\,(\mathbf{Z}\,\boldsymbol{\alpha}) = \mathbf{Z}^{(n-1)}\,\boldsymbol{\alpha}\ .$$

Using the chain rule of calculus, we obtain the next time derivative as

$$\frac{d^{(n)}}{dt^{(n)}}\,(\mathbf{Z}\,\boldsymbol{\alpha}) = \mathbf{Z}^{(n)}\,\boldsymbol{\alpha} + \mathbf{Z}^{(n-1)}\,\dot{\boldsymbol{\alpha}}\ .$$

Now, we want the function $\mathbf{Z}\,\boldsymbol{\alpha}$ to solve the inhomogeneous system (5.6):

$$\frac{d^{(n)}}{dt^{(n)}}\,(\mathbf{Z}\,\boldsymbol{\alpha}) = \left\{\sum_{i=0}^{n-1} \mathbf{A}_i(t)\,\mathbf{Z}^{(i)}(t)\right\}\boldsymbol{\alpha} + \boldsymbol{f}_{p_j}(t)\ .$$

On the other hand we know that the function $\mathbf{Z}$ "alone" solves the homogeneous system:

$$\mathbf{Z}^{(n)} = \left\{\sum_{i=0}^{n-1} \mathbf{A}_i(t)\,\mathbf{Z}^{(i)}(t)\right\}\ .$$

From the previous four relations the last line of the condition equations (5.17) easily follows.

The condition equations (5.17) may be written simply as

$$\tilde{\mathbf{Z}}\,\dot{\boldsymbol{\alpha}} = \boldsymbol{F}_{p_j} \ , \tag{5.18}$$

where the regular square matrix $\tilde{\mathbf{Z}}$ is defined as

$$\tilde{\mathbf{Z}} = \begin{pmatrix} \mathbf{Z} \\ \dot{\mathbf{Z}} \\ \ddot{\mathbf{Z}} \\ \cdots \\ \cdots \\ \mathbf{Z}^{(n-1)} \end{pmatrix} , \tag{5.19}$$

and the column array $\boldsymbol{F}_{p_j}$ is defined as

$$\boldsymbol{F}_{p_j} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \cdots \\ \cdots \\ \mathbf{0} \\ \boldsymbol{f}_{p_j} \end{pmatrix} . \tag{5.20}$$

Independently of the order n of the variational equation, a very simple differential equation system of first order is obtained for the parameter array $\boldsymbol{\alpha}$, which, as a matter of fact, may be solved by quadrature:

$$\boldsymbol{\alpha}(t) = \int_{t_0}^{t} \tilde{\mathbf{Z}}^{-1}(t')\,\boldsymbol{F}_{p_j}(t')\ dt' \ . \tag{5.21}$$

In mathematical textbooks the above deliberations are usually given for first-order systems of equations – what is sufficient from the point of view of pure mathematics, because every system of order $n > 1$ may be decomposed into one of first order. In numerical analysis, this step usually leads to a decrease of computational efficiency and to an increase of disk storage requirements.

The representation (5.21) is of crucial importance, if many (hundreds to thousands of) dynamical parameters have to be determined, because there are much more efficient tools available to solve integrals than differential equations (see Chapter 7).

The numerical solution of eqns. (5.21) may be optimized, if many variational equations referring to dynamical parameters have to be solved: Observe, e.g., that the matrix $\tilde{\mathbf{Z}}$ has to be inverted only once and that the matrix multiplication in the integrand has to be performed only over the last d elements of matrix $\boldsymbol{F}_{p_j}$ because only those elements are different from zero.

5.3 Variational Equations of the Two-Body Problem

The motion of two spherically symmetric bodies obeying Newton's law of universal gravitation is governed by the differential equations (4.1). According to the procedure outlined in section 5.1, the corresponding variational equations are obtained by taking the partial derivative of these equations w.r.t. one of the initial values (or w.r.t. one of the orbital elements, which are functions of these initial values). The result is

$$\begin{aligned} \ddot{\boldsymbol{z}} &= -\frac{\mu}{r^3} \begin{pmatrix} 1 - \dfrac{3\,r_1^2}{r^2} & , -\dfrac{3\,r_1\,r_2}{r^2} & , -\dfrac{3\,r_1\,r_2}{r^2} \\ -\dfrac{3\,r_1\,r_2}{r^2} & , 1 - \dfrac{3\,r_2^2}{r^2} & , -\dfrac{3\,r_2\,r_3}{r^2} \\ -\dfrac{3\,r_1\,r_3}{r^2} & , -\dfrac{3\,r_2\,r_3}{r^2} & , 1 - \dfrac{3\,r_3^2}{r^2} \end{pmatrix} \boldsymbol{z} \\ &\overset{\text{def}}{=} -\frac{\mu}{r^3} \left\{ \mathbf{E} - \frac{3}{r^2}\, \boldsymbol{r} \otimes \boldsymbol{r}^T \right\} \boldsymbol{z} \ . \end{aligned} \tag{5.22}$$

According to the definition (5.4) the column array $\boldsymbol{z}(t) \overset{\text{def}}{=} \left(\frac{\partial \boldsymbol{r}}{\partial p}\right)(t)$ is the partial derivative of the array $\boldsymbol{r}(t)$ w.r.t. one of the parameters p defining the initial values, $\boldsymbol{r} \otimes \boldsymbol{r}^T$ is the Cartesian product, or outer product, of the column array $\boldsymbol{r}$ with its transpose $\boldsymbol{r}^T$.

The variational equations obviously are a special case of eqns. (5.12), where the order of the system is $n = 2$ and its dimension $d = 3$, and where

$$\mathbf{A}_{00} \overset{\text{def}}{=} \mathbf{A}_0 = -\frac{\mu}{r^3} \left\{ \mathbf{E} - \frac{3}{r^2}\, \boldsymbol{r} \otimes \boldsymbol{r}^T \right\} \quad \text{and} \quad \mathbf{A}_{10} \overset{\text{def}}{=} \mathbf{A}_1 = \mathbf{0} \ . \tag{5.23}$$

The notations $\mathbf{A}_{00}$ and $\mathbf{A}_{10}$ will be used where necessary, to refer to the matrices accompanying the two-body problem. No dynamical parameters need to be considered in the case of the two-body problem.

The differential equations (4.1) of the two-body problem were solved in closed form in Chapter 4. In the remainder of this section we derive closed solutions of the variational equations (5.22).

According to eqns. (4.62, 4.64), and (4.66) the equations of motion (4.1) are solved in the quasi-inertial system by

$$\begin{aligned} \boldsymbol{r} &= \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \boldsymbol{r}_\Pi \\ \dot{\boldsymbol{r}} &= \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \dot{\boldsymbol{r}}_\Pi \ , \end{aligned} \tag{5.24}$$

where

$$\boldsymbol{r}_\Pi = \begin{pmatrix} r\,\cos v \\ r\,\sin v \\ 0 \end{pmatrix} \quad ; \qquad \dot{\boldsymbol{r}}_\Pi = \sqrt{\frac{\mu}{p}} \begin{pmatrix} -\sin v \\ e + \cos v \\ 0 \end{pmatrix} \tag{5.25}$$

are the component matrices of the position and velocity vector in the orbital system $\varPi$ as defined in Table 4.3 and illustrated by Figure 4.7.

As v, the celestial body's true anomaly, is defined in the same way for all types of orbits, we can state that the above equations hold for all orbit types (ellipses, parabolas, hyperbolas). $\boldsymbol{r}$ is the column matrix of Cartesian coordinates in the quasi-inertial system, $\boldsymbol{r}_{\varPi}$ the corresponding column matrix in the orbital system (orbital plane as fundamental plane, direction to the pericenter as first coordinate axis).

Subsequently, either of two alternative sets of orbital parameters will be used to define a particular solution of the two-body problem:

$$\{a, e, i, \varOmega, \omega, T_0\} \quad \text{or} \quad \{p, e, i, \varOmega, \omega, T_0\} \ , \tag{5.26}$$

where a is the semi-major axis, e the numerical eccentricity, p the semi-latus rectum, i the inclination w.r.t. reference plane, $\varOmega$ the longitude (or right ascension) of the ascending node, ω the argument of pericenter, and T_0 the time of pericenter passage.

The second of the sets (5.26) has the advantage to describe the orbit for all possible solutions, namely ellipse, parabola, and hyperbola (with the understanding that the eccentricity is fixed to $e = 1$ in the case of the parabola).

Other sets of six independent functions of the above orbital elements may be better suited for special cases (e.g., for low eccentricity elliptic orbits, or for low inclination orbits). The derivatives w.r.t. alternative sets of elements may be easily obtained by simple transformations of the derivatives provided subsequently.

The three Eulerian angles i, $\varOmega$, and ω only show up in the rotation matrices in eqns. (5.24) and (5.25), whereas the remaining three orbital elements are contained only in the two non-zero components of the state arrays $\boldsymbol{r}_{\varPi}$ and $\dot{\boldsymbol{r}}_{\varPi}$ in the orbital system.

The partial derivatives of the state vector w.r.t. the three Eulerian angles follow from eqns. (5.24):

$$\begin{aligned}
\frac{\partial}{\partial \varOmega}\{\boldsymbol{r}\} &= \frac{\partial}{\partial \varOmega}\big\{\mathbf{R}_3(-\varOmega)\big\}\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \boldsymbol{r}_{\varPi} = \frac{\partial}{\partial \varOmega}\{\mathbf{R}\}\ \boldsymbol{r}_{\varPi} \\
\frac{\partial}{\partial i}\{\boldsymbol{r}\} &= \mathbf{R}_3(-\varOmega)\ \frac{\partial}{\partial i}\big\{\mathbf{R}_1(-i)\big\}\ \mathbf{R}_3(-\omega)\ \boldsymbol{r}_{\varPi} = \frac{\partial}{\partial i}\{\mathbf{R}\}\ \boldsymbol{r}_{\varPi} \\
\frac{\partial}{\partial \omega}\{\boldsymbol{r}\} &= \mathbf{R}_3(-\varOmega)\ \mathbf{R}_1(-i)\ \frac{\partial}{\partial \omega}\big\{\mathbf{R}_3(-\omega)\big\}\ \boldsymbol{r}_{\varPi} = \frac{\partial}{\partial \omega}\{\mathbf{R}\}\ \boldsymbol{r}_{\varPi} \\
\frac{\partial}{\partial \varOmega}\{\dot{\boldsymbol{r}}\} &= \frac{\partial}{\partial \varOmega}\big\{\mathbf{R}_3(-\varOmega)\big\}\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \dot{\boldsymbol{r}}_{\varPi} = \frac{\partial}{\partial \varOmega}\{\mathbf{R}\}\ \dot{\boldsymbol{r}}_{\varPi} \\
\frac{\partial}{\partial i}\{\dot{\boldsymbol{r}}\} &= \mathbf{R}_3(-\varOmega)\ \frac{\partial}{\partial i}\big\{\mathbf{R}_1(-i)\big\}\ \mathbf{R}_3(-\omega)\ \dot{\boldsymbol{r}}_{\varPi} = \frac{\partial}{\partial i}\{\mathbf{R}\}\ \dot{\boldsymbol{r}}_{\varPi} \\
\frac{\partial}{\partial \omega}\{\dot{\boldsymbol{r}}\} &= \mathbf{R}_3(-\varOmega)\ \mathbf{R}_1(-i)\ \frac{\partial}{\partial \omega}\big\{\mathbf{R}_3(-\omega)\big\}\ \dot{\boldsymbol{r}}_{\varPi} = \frac{\partial}{\partial \omega}\{\mathbf{R}\}\ \dot{\boldsymbol{r}}_{\varPi}\ ,
\end{aligned} \tag{5.27}$$

where the product of the three rotation matrices was abbreviated as

$$\mathbf{R} \overset{\text{def}}{=} \mathbf{R}_3(-\varOmega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ . \tag{5.28}$$

It is a straight forward matter to verify that the rotation matrix $\mathbf{R}$ may be written explicitly as

$$\mathbf{R} = \begin{pmatrix} \cos\varOmega\cos\omega - \sin\varOmega\cos i\sin\omega\ , & -\cos\varOmega\sin\omega - \sin\varOmega\cos i\cos\omega\ , & \ldots \\ \sin\varOmega\cos\omega + \cos\varOmega\cos i\sin\omega\ , & -\sin\varOmega\sin\omega + \cos\varOmega\cos i\cos\omega\ , & \ldots \\ \sin i\sin\omega\ , & \sin i\cos\omega\ , & \ldots \end{pmatrix} . \tag{5.29}$$

Note that only the first two columns of matrix $\mathbf{R}$ (and of its partial derivatives below) are needed, because the third components of the state vector is zero in the orbital system.

The partial derivatives of matrix $\mathbf{R}$ w.r.t. the three Eulerian angles may then be written explicitly as

$$\frac{\partial}{\partial \Omega}\{\mathbf{R}\} =$$

$$\begin{pmatrix} -\sin\Omega\cos\omega - \cos\Omega\cos i\sin\omega & , +\sin\Omega\sin\omega - \cos\Omega\cos i\cos\omega & , \dots \\ +\cos\Omega\cos\omega - \sin\Omega\cos i\sin\omega & , -\cos\Omega\sin\omega - \sin\Omega\cos i\cos\omega & , \dots \\ 0 & , \quad 0 & , \dots \end{pmatrix} ,$$

$$\frac{\partial}{\partial i}\{\mathbf{R}\} = \begin{pmatrix} +\sin\Omega\sin i\sin\omega & , +\sin\Omega\sin i\cos\omega & , \dots \\ -\cos\Omega\sin i\sin\omega & , -\cos\Omega\sin i\cos\omega & , \dots \\ \cos i\sin\omega & , \quad \cos i\cos\omega & , \dots \end{pmatrix} ,$$

$$\frac{\partial}{\partial \omega}\{\mathbf{R}\} =$$

$$\begin{pmatrix} -\cos\Omega\sin\omega - \sin\Omega\cos i\cos\omega & , -\cos\Omega\cos\omega + \sin\Omega\cos i\sin\omega & , \dots \\ -\sin\Omega\sin\omega + \cos\Omega\cos i\cos\omega & , -\sin\Omega\cos\omega - \cos\Omega\cos i\sin\omega & , \dots \\ +\sin i\cos\omega & , \quad -\sin i\sin\omega & , \dots \end{pmatrix} . \tag{5.30}$$

As the angles Ω, i, and ω are constants of integration of the two-body problem, the matrix $\mathbf{R}$ and its partial derivatives are matrices with constant elements, as well. Because the components $\boldsymbol{r}_\Pi$ and $\dot{\boldsymbol{r}}_\Pi$ are periodic functions of time in the case of the elliptic motion, the partial derivatives of the state vector of an elliptic orbit w.r.t. the three Eulerian angles are periodic functions, as well.

This leaves us with the partial derivatives w.r.t. the four elements a, p, e, and T_0 (where only three are independent). Let therefore

$$\tilde{p} \in \{a, p, e, T_0\} \ . \tag{5.31}$$

Because the rotation matrix $\mathbf{R}$ does not depend on the three elements considered now, we may conclude from eqns. (5.24):

$$\begin{aligned} \frac{\partial}{\partial \tilde{p}}\{\boldsymbol{r}\} &= \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \frac{\partial}{\partial \tilde{p}}\{\boldsymbol{r}_\Pi\} \\ \frac{\partial}{\partial \tilde{p}}\{\dot{\boldsymbol{r}}\} &= \mathbf{R}_3(-\Omega)\ \mathbf{R}_1(-i)\ \mathbf{R}_3(-\omega)\ \frac{\partial}{\partial \tilde{p}}\{\dot{\boldsymbol{r}}_\Pi\} \ . \end{aligned} \tag{5.32}$$

From here onwards we have to distinguish between the three types of orbits, namely ellipses, parabolas, and hyperbolas.

5.3.1 Elliptic Orbits

In the case of the ellipse the semi-latus rectum p, the numerical eccentricity e, and the semi-major axis a are related by:

$$p = a\left(1 - e^2\right) , \tag{5.33}$$

which means that we may derive the formulas for the partial derivatives for two of the three elements and give the partial derivative w.r.t. the third element as a function of the two others. We prefer to use a and e as independent elements. By virtue of the above equation defining the semi-latus rectum of the ellipse, the operator for the partial derivative w.r.t. the parameter p is given by:

$$\frac{\partial}{\partial p} = \frac{\partial a}{\partial p}\,\frac{\partial}{\partial a} + \frac{\partial e}{\partial p}\,\frac{\partial}{\partial e} = \frac{1}{1-e^2}\,\frac{\partial}{\partial a} - \frac{1}{2\,a\,e}\,\frac{\partial}{\partial e} . \tag{5.34}$$

From eqns. (5.25) and the formulas of the two-body problem we see that the components of the state vector (in the orbital system) have the following structure:

$$\begin{aligned} \boldsymbol{r}_\Pi &= \boldsymbol{r}_\Pi\big(v(a,e,T_0); a, e\big) \\ \dot{\boldsymbol{r}}_\Pi &= \dot{\boldsymbol{r}}_\Pi\big(v(a,e,T_0); a, e\big) . \end{aligned} \tag{5.35}$$

Consequently, the partials w.r.t. the three elements a, e, and T_0 may computed as:

$$\begin{aligned} \frac{\partial}{\partial a}\{\boldsymbol{r}_\Pi\} &= \{\boldsymbol{r}_\Pi\}_a + \frac{\partial}{\partial v}\{\boldsymbol{r}_\Pi\}\,\frac{\partial v}{\partial a} \\ \frac{\partial}{\partial e}\{\boldsymbol{r}_\Pi\} &= \{\boldsymbol{r}_\Pi\}_e + \frac{\partial}{\partial v}\{\boldsymbol{r}_\Pi\}\,\frac{\partial v}{\partial e} \\ \frac{\partial}{\partial T_0}\{\boldsymbol{r}_\Pi\} &= \frac{\partial}{\partial v}\{\boldsymbol{r}_\Pi\}\,\frac{\partial v}{\partial T_0} \\ \frac{\partial}{\partial a}\{\dot{\boldsymbol{r}}_\Pi\} &= \{\dot{\boldsymbol{r}}_\Pi\}_a + \frac{\partial}{\partial v}\{\dot{\boldsymbol{r}}_\Pi\}\,\frac{\partial v}{\partial a} \\ \frac{\partial}{\partial e}\{\dot{\boldsymbol{r}}_\Pi\} &= \{\dot{\boldsymbol{r}}_\Pi\}_e + \frac{\partial}{\partial v}\{\dot{\boldsymbol{r}}_\Pi\}\,\frac{\partial v}{\partial e} \\ \frac{\partial}{\partial T_0}\{\dot{\boldsymbol{r}}_\Pi\} &= \frac{\partial}{\partial v}\{\dot{\boldsymbol{r}}_\Pi\}\,\frac{\partial v}{\partial T_0} , \end{aligned} \tag{5.36}$$

where $\{\boldsymbol{r}_\Pi\}_a$ and $\{\boldsymbol{r}_\Pi\}_e$ designate the partial derivatives of the coordinates in the orbital system w.r.t. a and e, ignoring, however, the dependency of the true anomaly v on the two elements.

It is a straight forward matter to compute these partial derivatives starting from eqns. (5.25):

$$\begin{aligned}
\{\boldsymbol{r}_{\Pi}\}_a &= \frac{r}{a}\begin{pmatrix}\cos v\\ \sin v\\ 0\end{pmatrix} &;\quad \{\dot{\boldsymbol{r}}_{\Pi}\}_a &= -\frac{1}{2\,a}\sqrt{\frac{\mu}{p}}\begin{pmatrix}-\sin v\\ e+\cos v\\ 0\end{pmatrix}\\
\{\boldsymbol{r}_{\Pi}\}_e &= -\,a\,\frac{r}{p}\left(\frac{r}{p}+e\right)\begin{pmatrix}\cos v\\ \sin v\\ 0\end{pmatrix} &;\quad \{\dot{\boldsymbol{r}}_{\Pi}\}_e &= \sqrt{\frac{\mu}{p}}\,\frac{ae}{p}\begin{pmatrix}-\sin v\\ \frac{p}{a\,e}+e+\cos v\\ 0\end{pmatrix}\\
\frac{\partial}{\partial v}\{\boldsymbol{r}_{\Pi}\} &= \frac{r^2}{p}\begin{pmatrix}-\sin v\\ e+\cos v\\ 0\end{pmatrix} &;\quad \frac{\partial}{\partial v}\{\dot{\boldsymbol{r}}_{\Pi}\} &= -\sqrt{\frac{\mu}{p}}\begin{pmatrix}\cos v\\ \sin v\\ 0\end{pmatrix}\ .
\end{aligned} \tag{5.37}$$

In order to calculate the partial derivatives of the true anomly w.r.t. the elements a, e, and T_0, we need the transformation between the true and the eccentric anomalies v and E, Kepler's equation, and the equation defining the mean motion $\sigma(t)$ (see Table 4.2):

$$\begin{aligned}
\tan\frac{v}{2} &= \sqrt{\frac{1+e}{1-e}}\,\tan\frac{E}{2}\\
E &= \sigma(t)+e\sin E\\
\sigma(t) &= \sqrt{\frac{\mu}{a^3}}\,(t-T_0)\ .
\end{aligned} \tag{5.38}$$

Alternatively, the equations (4.55) might be used. The advantage of using these equations resides in the elimination of the "auxiliary" angles E (elliptic motion) and F (hyperbolic motion). The disadvantage has to be seen in the complexity of eqns. (4.55).

The structure of eqns. (5.38) is obviously as follows:

$$\begin{aligned}
v &= v(e,E)\\
E &= E(\sigma,e)\\
\sigma &= \sigma(a,T_0)\ .
\end{aligned} \tag{5.39}$$

In order to reduce the above formulas to the essential content, the time dependency (which would enter into all equations) was left out.

Making use of this structure we may compute the partial derivatives of the true anomaly w.r.t. the elements a, e, and T_0 systematically as follows:

$$\begin{aligned}\frac{\partial v}{\partial a} &= \frac{\partial v}{\partial E}\,\frac{\partial E}{\partial \sigma}\,\frac{\partial \sigma}{\partial a} \\ \frac{\partial v}{\partial e} &= \{v\}_e + \frac{\partial v}{\partial E}\,\frac{\partial E}{\partial e} \\ \frac{\partial v}{\partial T_0} &= \frac{\partial v}{\partial E}\,\frac{\partial E}{\partial \sigma}\,\frac{\partial \sigma}{\partial T_0}\ ,\end{aligned} \tag{5.40}$$

where v_e designates the partial derivative of the true anomaly v w.r.t. e, ignoring the dependency of E on e.

All that remains to be done is the calculation of the partial derivatives on the right-hand side of the above equation.

From the first of equations (5.38) we obtain the partial derivative of the true anomaly w.r.t. the eccentric anomaly:

$$\begin{aligned}\frac{1}{2}\,\frac{1}{\cos^2\frac{v}{2}}\,\frac{\partial v}{\partial E} &= \sqrt{\frac{1+e}{1-e}}\,\frac{1}{2}\,\frac{1}{\cos^2\frac{E}{2}} \\ \frac{\partial v}{\partial E} &= \sqrt{\frac{1+e}{1-e}}\,\frac{1+\cos v}{1+\cos E} \\ &= \frac{1}{\sqrt{1-e^2}}\,\frac{p}{r}\ .\end{aligned} \tag{5.41}$$

The first two lines of the above derivation are straight forward, for the third line one needs the equation $r\,\cos v = a\,(\cos E - e)$ (see eqns. (4.62)).

The partial derivative of the eccentric anomaly E w.r.t. the mean anomaly σ follows from Kepler's equation (second of eqns. (5.38)):

$$\frac{\partial E}{\partial \sigma} = \frac{a}{r}\ . \tag{5.42}$$

In order to obtain the partial derivative of the true anomaly v w.r.t. the semi-major axis a we need, according to the first of eqns. (5.40), in addition the partial derivative of the mean anomaly w.r.t. the semi-major axis. This relation follows in turn from the third of eqns. (5.38):

$$\frac{\partial \sigma}{\partial a} = -\,\frac{3}{2a}\,\sigma(t)\ . \tag{5.43}$$

The partial derivative of the true anomaly v w.r.t. the eccentricity e (without considering the dependency of the eccentric anomaly E on the eccentricity e) follows first from taking the partial of eqn. (5.38) w.r.t. e:

$$\begin{aligned}
\frac{1}{2}\,\frac{1}{\cos^2\frac{v}{2}}\,\{v\}_e &= \frac{\partial}{\partial e}\left\{\sqrt{\frac{1+e}{1-e}}\right\}\tan\frac{E}{2} \\
\{v\}_e &= \sqrt{\frac{1+e}{1-e}}\,\frac{2}{1-e^2}\,\tan\frac{E}{2}\,\cos^2\frac{v}{2} \\
&= \sqrt{\frac{1+e}{1-e}}\,\frac{1}{1-e^2}\,\sin E\,\frac{1+\cos v}{1+\cos E} \\
&= \frac{1}{\left(1-e^2\right)^2}\,\frac{r}{a}\,\sin v\,\left(1+e\,\cos v\right) \\
&= \frac{1}{\left(1-e^2\right)^2}\,\frac{p}{a}\,\sin v \\
&= \frac{\sin v}{1-e^2}\;,
\end{aligned} \tag{5.44}$$

then by taking the partial derivative of the eccentric anomaly w.r.t. the eccentricity:

$$\frac{\partial E}{\partial e} = \frac{a}{r}\,\sin E = \frac{\sin v}{\sqrt{1-e^2}}\;. \tag{5.45}$$

As we have already calculated the partial derivative of E w.r.t. σ, the preceding two relations allow us to calculate the partial derivative of the true anomaly v w.r.t. the eccentricity e with the second of equations (5.40). The partial derivative of the true anomaly w.r.t. T_0 follows from the third of the same set of equations, where the partial derivative of the mean anomaly σ w.r.t. the time of pericenter is

$$\frac{\partial \sigma}{\partial T_0} = -\sqrt{\frac{\mu}{a^3}}\;. \tag{5.46}$$

We are now in a position to calculate – in closed form ("analytically") – the partial derivatives of the two-body orbit (and its velocity) w.r.t. all six elements a, e, i, Ω, ω, and T_0 as a function of time t.

The six functions $\boldsymbol{z}_1 \overset{\text{def}}{=} \frac{\partial \boldsymbol{r}}{\partial a}(t)\,, \boldsymbol{z}_2 \overset{\text{def}}{=} \frac{\partial \boldsymbol{r}}{\partial e}(t)\,, \ldots, \boldsymbol{z}_6 \overset{\text{def}}{=} \frac{\partial \boldsymbol{r}}{\partial T_0}(t)$ form a complete system of solutions of the homogeneous variational equations (5.22) accompanying the two-body problem in the case of elliptic orbits.

5.3.2 Parabolic Orbits

Parabolic orbits are best described by the second set of orbital parameters (5.26). As the numerical eccentricity is constrained to $e = 1$, we have to deal with the five orbital parameters $\{p, i, \Omega, \omega, T_0\}$.

The partial derivatives w.r.t. the three Eulerian angles i, Ω, and ω are the same for all three types of orbits. Therefore, we only have to derive the partial derivatives w.r.t. the semi-latus rectum p and the time T_0 of pericenter.

In order to do that, we first have to transform the derivatives (5.37) into derivatives w.r.t. the semi-latus rectum p and the true anomaly v (where we merely have to set $e = 1$ in the latter case):

$$\begin{aligned}
\{\boldsymbol{r}_\Pi\}_p &= \frac{r}{p}\begin{pmatrix}\cos v\\ \sin v\\ 0\end{pmatrix} \quad ; \quad \{\dot{\boldsymbol{r}}_\Pi\}_p = -\frac{1}{2}\sqrt{\frac{\mu}{p^3}}\begin{pmatrix}-\sin v\\ 1+\cos v\\ 0\end{pmatrix}\\
\{\boldsymbol{r}_\Pi\}_v &= \frac{r^2}{p}\begin{pmatrix}-\sin v\\ 1+\cos v\\ 0\end{pmatrix} \quad ; \quad \{\dot{\boldsymbol{r}}_\Pi\}_v = -\sqrt{\frac{\mu}{p}}\begin{pmatrix}\cos v\\ \sin v\\ 0\end{pmatrix} \; .
\end{aligned} \tag{5.47}$$

According to Table 4.2 the true anomaly v may be computed without any transformations as a function of time

$$\tan\frac{v}{2} + \frac{1}{3}\tan^3\frac{v}{2} = 2\sqrt{\frac{\mu}{p^3}}\,(t - T_0) \; . \tag{5.48}$$

The true anomaly v is a function of the semi-latus p of the parabola and of the time T_0 of pericenter passage (and of course of the time t).

The partial derivatives w.r.t. p and T_0 are formed according to the same pattern as in the case of the ellipse. The formula are simpler because no auxiliary angle has to be introduced. Equations (5.36) have to be replaced by the relations

$$\begin{aligned}
\frac{\partial}{\partial p}\{\boldsymbol{r}_\Pi\} &= \{\boldsymbol{r}_\Pi\}_p + \frac{\partial}{\partial v}\{\boldsymbol{r}_\Pi\}\,\frac{\partial v}{\partial p}\\
\frac{\partial}{\partial T_0}\{\boldsymbol{r}_\Pi\} &= \frac{\partial}{\partial v}\{\boldsymbol{r}_\Pi\}\,\frac{\partial v}{\partial T_0}\\
\frac{\partial}{\partial p}\{\dot{\boldsymbol{r}}_\Pi\} &= \{\dot{\boldsymbol{r}}_\Pi\}_p + \frac{\partial}{\partial v}\{\dot{\boldsymbol{r}}_\Pi\}\,\frac{\partial v}{\partial p}\\
\frac{\partial}{\partial T_0}\{\dot{\boldsymbol{r}}_\Pi\} &= \frac{\partial}{\partial v}\{\dot{\boldsymbol{r}}_\Pi\}\,\frac{\partial v}{\partial T_0} \; .
\end{aligned} \tag{5.49}$$

All that remains to be done is the calculation of the partial derivatives of the true anomaly w.r.t. the elements p and T_0, by taking partial derivative of eqn. (5.48) w.r.t. the corresponding element

$$\begin{aligned}
\frac{\partial v}{\partial p} &= -\frac{3}{2}\frac{p^2}{r^2}\sqrt{\frac{\mu}{p^5}}\,(t - T_0)\\
\frac{\partial v}{\partial T_0} &= -\frac{p^2}{r^2}\sqrt{\frac{\mu}{p^3}} \; ,
\end{aligned} \tag{5.50}$$

where use was made of the (elementary) relations

$$\frac{d}{dv}\left\{\tan\frac{v}{2}\right\} = \frac{1}{2}\,\frac{1}{\cos^2\frac{v}{2}} = \frac{1}{1+\cos v} = \frac{r}{p}\ . \tag{5.51}$$

Observe that the latter equality only holds for parabolas, whereas the first hold for all conic sections.

With the equations derived in this paragraph we are in a position to calculate the partial derivatives of the state vector w.r.t. all five orbital elements in the case of the parabola.

In analogy to the elliptic motion only the partial derivative w.r.t. one orbital element, namely p, grows linearly with the time t. The other partial derivatives are of course not periodic (there is no period in the case of parabolic motion), but their absolute values are constrained by $\sin v$ or $\cos v$.

5.3.3 Hyperbolic Orbits

The partial derivatives w.r.t. the three Eulerian angles obey the formulas (5.27), which are (as pointed out previously) independent of the shape of the orbit.

The partial derivatives w.r.t. the elements a, e, and T_0 are calculated according to the same pattern as in the case of the ellipse. The only difference resides in the facts that the eccentric anomaly E has to be replaced by the hyperbolic analogue F, and that Kepler's equation has to be replaced by the corresponding equation in the case of hyperbolic motion (see Table 4.2). The eqns. (5.38) thus have to be replaced by the following set of equations (where only the first two are actually different from the set (5.38)):

$$\begin{aligned}
\tan\frac{v}{2} &= \sqrt{\frac{e+1}{e-1}}\,\tanh\frac{F}{2}\\
F &= e\,\sinh F\ -\ \sigma(t)\\
\sigma(t) &= \sqrt{\frac{\mu}{a^3}}\,(t-T_0)\ .
\end{aligned} \tag{5.52}$$

Equations (5.36) and (5.37) may be taken over without change from the elliptic motion, whereas eqns. (5.40) have to be modified as follows:

$$\begin{aligned}
\frac{\partial v}{\partial a} &= \frac{\partial v}{\partial F}\,\frac{\partial F}{\partial \sigma}\,\frac{\partial \sigma}{\partial a}\\
\frac{\partial v}{\partial e} &= \{v\}_e + \frac{\partial v}{\partial F}\,\frac{\partial F}{\partial e}\\
\frac{\partial v}{\partial T_0} &= \frac{\partial v}{\partial F}\,\frac{\partial F}{\partial \sigma}\,\frac{\partial \sigma}{\partial T_0}\ .
\end{aligned} \tag{5.53}$$

The three partial derivatives needed to calculate all partial derivatives of the state vector w.r.t. the elements a, e, and T_0 are:

$$\begin{aligned}
\frac{\partial v}{\partial F} &= \sqrt{\frac{e+1}{e-1}}\,\frac{1+\cos v}{1+\cosh F} = \sqrt{e^2-1}\,\frac{a}{r} \\
\{v\}_e &= -\sqrt{\frac{e+1}{e-1}}\,\frac{1+\cos v}{e^2-1}\,\tanh\frac{F}{2} = -\frac{\sin v}{e^2-1} \\
\frac{\partial F}{\partial \sigma} &= \frac{1}{e\,\cosh F - 1} = \frac{a}{r}\ ,
\end{aligned} \tag{5.54}$$

where the first of eqns. (5.52) and the relation

$$r = \frac{a\left(e^2-1\right)}{1+e\,\cos v} = a\left(e\,\cosh F - 1\right) \tag{5.55}$$

were used to derive the results (5.54).

5.3.4 Summary and Examples

Analytical solutions (solutions in mathematically closed form) of the variational equations (5.22) were derived for the three types of two-body orbits in section 5.3. The orbital elements rather than the components of the initial state vectors were used to parametrize the problem. Complete sets of six solutions for the solution of the variational equations were given for elliptic and hyperbolic orbits. (For obvious reasons only five partial derivatives were provided in the case of the parabola.)

Figure 5.1 shows the partial derivatives of an unperturbed two-body orbit w.r.t. four out of the six elements (namely for a, e, i, and ω). The examples refer to (hypothetical) minor planets with revolution periods P of about four years (about a third of the revolution period $P_{♃}$ of Jupiter) with moderate eccentricity and inclination (the precise values do not matter in our context).

When interpreting Figures 5.1 we should keep in mind that the product of a partial derivative w.r.t. an orbital element with the difference in the corresponding orbital element equals the following differences of solutions of the two-body problem

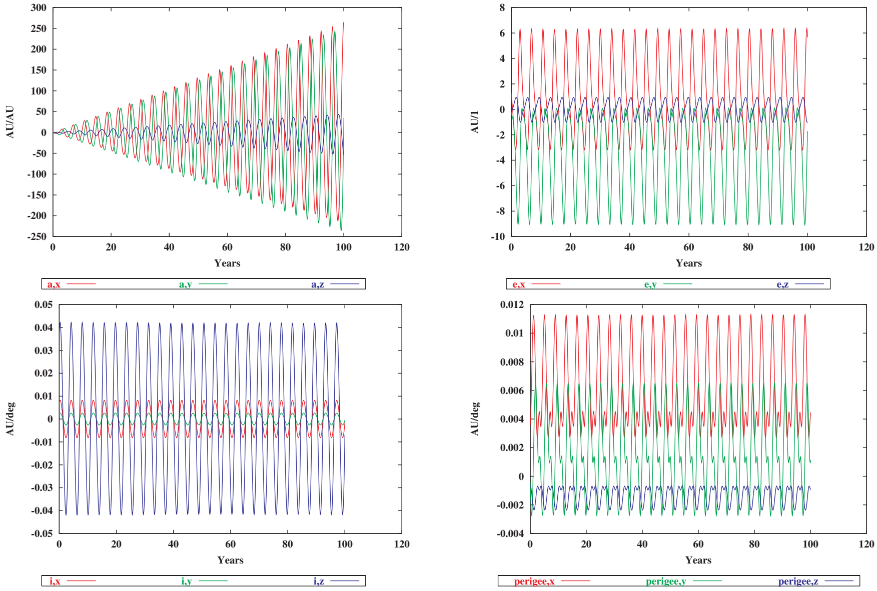


Fig. 5.1. Partial derivatives of a two-body orbit w.r.t. semi-major axis a and eccentricity e (first row), inclination i, and argument of pericenter ω (second row) over 100 years ($P = 0.326\ P_{♃}$, $e = 0.10$, $i = 11.58^\circ$, $\Omega = 107.6^\circ$)

$$
\begin{aligned}
\frac{\partial \boldsymbol{r}}{\partial a}(t)\ \Delta a \quad &= \boldsymbol{r}(t; a + \Delta a, e, i, \Omega, \omega, T_0) - \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0) \\
\frac{\partial \boldsymbol{r}}{\partial e}(t)\ \Delta e \quad &= \boldsymbol{r}(t; a, e + \Delta e, i, \Omega, \omega, T_0) - \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0) \\
\frac{\partial \boldsymbol{r}}{\partial i}(t)\ \Delta i \quad &= \boldsymbol{r}(t; a, e, i + \Delta i, \Omega, \omega, T_0) - \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0) \\
\frac{\partial \boldsymbol{r}}{\partial \Omega}(t)\ \Delta \Omega &= \boldsymbol{r}(t; a, e, i, \Omega + \Delta \Omega, \omega, T_0) - \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0) \\
\frac{\partial \boldsymbol{r}}{\partial \omega}(t)\ \Delta \omega \quad &= \boldsymbol{r}(t; a, e, i, \Omega, \omega + \Delta \omega, T_0) - \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0) \\
\frac{\partial \boldsymbol{r}}{\partial T_0}(t)\ \Delta T_0 &= \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0 + \Delta T_0) - \boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0)\ ,
\end{aligned}
\tag{5.56}
$$

provided the differences Δa, Δe, Δi, $\Delta \Omega$, $\Delta \omega$, and ΔT_0 in the orbital elements are infinitesimally small.

With this understanding Figures 5.1 may be easily interpreted: They illustrate the development of the difference of two orbits, which were infinitesimally close to each other at the initial epoch t_0. If the semi-major axis of the reference orbit $\boldsymbol{r}(t; a, e, i, \Omega, \omega, T_0)$ is changed by a small amount at time

t_0, Figure 5.1 tells, that after one hundred years (corresponding to about 25 revolutions) the resulting effect on the orbit is amplified by a factor of about 250, whereas only periodic variations with small amplitudes are expected when changing the other five elements.

The remarkable difference in the signature of the partial derivative of a reference orbit w.r.t. the semi-major axis a when compared to the partials w.r.t. one of the other orbital elements is explained by the fact that a defines the mean motion via Kepler's third law $n^2 a^3 = \mu$ (see eqn. (4.41)). The oscillations in Figure 5.1 with linearly growing amplitudes are explained as the difference of two position vectors corresponding to orbits with slightly different mean motions (the difference of mean anomalies grows linearly with time t).

Figures 5.1 are (of course) characteristic for the partial derivatives of the two-body motion, but also for the perturbed two-body motion – provided the time interval considered is not too long and the perturbations are small compared to the main term. If two perturbed orbits, which were infinitesimally close at t_0, evolve according to the pattern of Figure 5.1 (top, left), this merely implies that the two semi-major axes (and consequently the corresponding mean motions) slightly differ. Figures with the signature of Figure 5.1 are often incorrectly interpreted.

The above figures were generated with program PLASYS (see Chapter II- 10 of Part III) and not with the analytical formulas developed above. The results are, however, undistinguishable from the analytical solutions developed here.

5.4 Variational Equations Associated with One Trajectory

In Chapter 3 the equations of motion (3.21) for a celestial body of negligible mass (in the planetary system) were written as:

$$\ddot{\boldsymbol{r}} = -\,k^2\, m_0\, \frac{\boldsymbol{r}}{r^3} \;-\; k^2 \sum_{j=1}^{n} m_j \left\{ \frac{\boldsymbol{r} - \boldsymbol{r}_j}{|\boldsymbol{r} - \boldsymbol{r}_j|^3} + \frac{\boldsymbol{r}_j}{r_j^3} \right\} \;. \tag{5.57}$$

Following the procedure outlined in section 5.1 we take the partial derivative of the above equation w.r.t. one of the parameters defining the initial state vector at time t_0, or with respect to one of the dynamical parameters, i.e., the masses m_j of the planets. As there are no velocity-dependent forces in this case, the result may be written in the form

$$\ddot{\boldsymbol{z}}_p = \mathbf{A}_0\, \boldsymbol{z}_p + \boldsymbol{f}_p \;, \tag{5.58}$$

where

$$\begin{aligned}\mathbf{A}_0 &= -\frac{k^2}{r^3}\left[\mathbf{E} - \frac{3}{r^2}\,\boldsymbol{r}\otimes\boldsymbol{r}^T\right]\\ &\quad - k^2\sum_{j=1}^{n}\frac{m_j}{|\boldsymbol{r}-\boldsymbol{r}_j|^3}\left[\mathbf{E} - \frac{3}{|\boldsymbol{r}-\boldsymbol{r}_j|^2}\,(\boldsymbol{r}-\boldsymbol{r}_j)\otimes(\boldsymbol{r}-\boldsymbol{r}_j)^T\right] \qquad (5.59)\\ &\overset{\text{def}}{=} \mathbf{A}_{00} + \varDelta\mathbf{A}_0\end{aligned}$$

and

$$\boldsymbol{f}_p = \begin{cases} -\,k^2\left\{\dfrac{\boldsymbol{r}-\boldsymbol{r}_j}{|\boldsymbol{r}-\boldsymbol{r}_j|^3} + \dfrac{\boldsymbol{r}_j}{r_j^3}\right\} & ; \quad \text{for } p = m_j \\ \mathbf{0} & ; \quad \text{for } p \in \{a_0, e_0, i_0, \varOmega_0, \omega_0, T_{00}\}\ . \end{cases} \qquad (5.60)$$

Observe that the initial values are assumed to be defined by the initial osculating elements $a_0, e_0, i_0, \varOmega_0, \omega_0, T_{00}$, referring to the initial epoch t_0.

One easily sees from eqn. (5.59) that the matrix $\mathbf{A}_0$ is composed of the two-body constituent $\mathbf{A}_{00}$ and the (small) contribution $\varDelta\mathbf{A}_0$ due to the perturbations.

The initial conditions associated with the variational equation (5.58) are defined by

$$\begin{aligned} \boldsymbol{z}_p(t_0) &= \begin{cases} \dfrac{\partial \boldsymbol{r}_0}{\partial p} & \text{for } \ p \in \{a_0, e_0, i_0, \varOmega_0, \omega_0, T_{00}\} \\ \mathbf{0} & \text{for } \ p \in \{m_1, m_2, \ldots, m_n\} \end{cases} \\ \dot{\boldsymbol{z}}_p(t_0) &= \begin{cases} \dfrac{\partial \dot{\boldsymbol{r}}_0}{\partial p} & \text{for } \ p \in \{a_0, e_0, i_0, \varOmega_0, \omega_0, T_{00}\} \\ \mathbf{0} & \text{for } \ p \in \{m_1, m_2, \ldots, m_n\}\ . \end{cases} \end{aligned} \qquad (5.61)$$

The partial derivatives $\frac{\partial \boldsymbol{r}_0}{\partial p}$ and $\frac{\partial \dot{\boldsymbol{r}}_0}{\partial p}$ of the initial state vector have to be calculated according to the formulas of the two-body problem developed previously.

The variational equations in program PLASYS are solved exactly according to the procedure outlined here, where the equations are simultaneously integrated with the primary equations (5.57).

Figure 5.2 illustrates the solution of the variational equations in the presence of the perturbations by Jupiter, Saturn, Uranus, Neptune, and Pluto (where the dominating influence is due to Jupiter and Saturn) over a time interval of 1000 years.

The variational equations correspond to two orbits close to the (3:1)-commensurability with Jupiter. The revolution period of the first orbit is $P = 0.3263\bar{3}\,P_{♃}$, of the second orbit it is $P = 0.333\bar{3}\,P_{♃}$. The solutions of the variational equations (elements a, e, and i) are contained in the first ($P = 0.3263\bar{3}\,P_{♃}$) and second ($P = 0.333\bar{3}\,P_{♃}$) column in Figure 5.2.

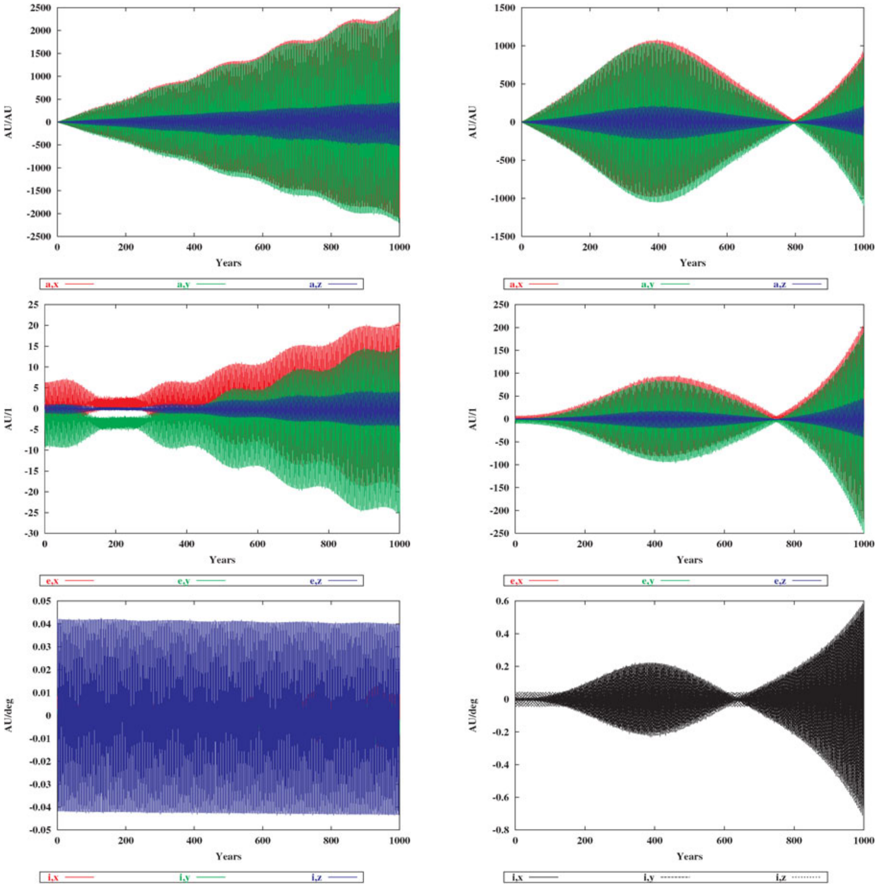


Fig. 5.2. Partial derivatives of a perturbed orbit (by Jupiter, Saturn, Uranus, Neptune, and Pluto) w.r.t. semi-major axis a, eccentricity e, and inclination i over 1000 years ($P = 0.326\bar{3}\,P_{♃}$ (left), $P = 0.333\bar{3}\,P_{♃}$ (right) , $e = 0.10$, $i = 11.58°$, $\Omega = 107.6°$, $\tilde{\omega} - \tilde{\omega_{♃}} = 90°$, $T_0 = T_{0♃}$, $t_0 = 2000$, Jan 1.0; $P_{♃}$ is Jupiter's orbital period, $\tilde{\omega}$, $\tilde{\omega}_{♃}$ the test particle's and Jupter's perihelion longitudes)

The initial conditions of the primary equations corresponding to the left column in Figure 5.2 are identical as those in Figure 5.1, where the variational equations of the two-body problem are shown. For the first 100 to 200 years (corresponding to 25 to 50 revolutions), the unperturbed and the perturbed solutions of the variational equations are very similar. In view of the fact that the perturbations are small compared to the main term, this result could be expected (it was anticipated in the previous section). It implies, that in parameter estimation procedures covering time intervals of only few revolutions, it may be sufficient to approximate the partial derivatives of the orbit w.r.t. the initial osculating elements by the two-body approximation.

Note that in the case of the perturbed motion, after an initial time span of a few dozen revolutions, the signature of the partial derivatives may deviate significantly (in the second column of Figures 5.2 even dramatically) from the signature observed for the two-body motion in Figure 5.1. Obviously, a small change in the osculating elements other than a at t_0 also influences the mean motion of the minor planet.

The equations of motion for an artificial Earth satellite were derived in Chapter 3, eqns. (3.143) and (3.144). One easily verifies that their structure is

$$\ddot{\boldsymbol{r}} = -GM\,\frac{\boldsymbol{r}}{r^3} + \delta\boldsymbol{f}\ , \tag{5.62}$$

where the perturbation term may be much more complicated than in the case of the motion in the planetary system. The structure of the variational equations associated with eqns. (5.62) is

$$\ddot{\boldsymbol{z}} = \mathbf{A}_{00}\,\boldsymbol{z} + \delta\mathbf{A}_0\,\boldsymbol{z} + \mathbf{A}_1\,\dot{\boldsymbol{z}} + \delta\boldsymbol{f}_p\ , \tag{5.63}$$

where $\mathbf{A}_{00}$ is the matrix of the two-body problem (see eqns. (5.23)). The concrete form of the other matrices depend on the concrete orbit model used and on the particular parameter.

In satellite geodesy the determination of dynamical parameters plays a much more important role than in problems related to the planetary system. If the gravity field of the Earth is determined from the orbital motion of close Earth satellites, thousands of parameters C_{ik}, S_{ik} have to be solved for, whereas only relatively few osculating elements have to be determined (the actual number depends on the length of the satellite arcs analyzed).

Note, that the structure of the variational equations describing the satellite motion is in essence the same as the structure of the variational equations (5.58) associated with the motion of a minor planet or comet in the planetary system. An important difference resides in the fact that matrix $\mathbf{A}_1 \neq \mathbf{0}$ for low Earth orbiters (LEOs).

5.5 Variational Equations Associated with the N-Body Problem

Equations (3.18) are the equations of motion of the entire planetary system. They were derived on the basis of the Newton-Euler equations of motion, assuming point masses for all celestial bodies involved. A particular solution of these equations is defined by the following initial value problem (the argument of latitude u_0 is used subsequently instead of the time T_0 of pericenter passage):

$$\begin{aligned}
\ddot{\boldsymbol{r}}_i \quad &= -k^2\,(m_0+m_i)\,\frac{\boldsymbol{r}_i}{r_i^3} - k^2 \sum_{j=1,j\neq i}^{n} m_j \left[\frac{\boldsymbol{r}_i-\boldsymbol{r}_j}{|\boldsymbol{r}_i-\boldsymbol{r}_j|^3}+\frac{\boldsymbol{r}_j}{r_j^3}\right] \\
\boldsymbol{r}_i(t_0) &\stackrel{\text{def}}{=} \boldsymbol{r}_{i0}(a_{i0},e_{i0},i_{i0},\Omega_{i0},\omega_{i0},u_{i0}) \\
\dot{\boldsymbol{r}}_i(t_0) &\stackrel{\text{def}}{=} \dot{\boldsymbol{r}}_{i0}(a_{i0},e_{i0},i_{i0},\Omega_{i0},\omega_{i0},u_{i0}) \\
& \qquad\qquad\qquad\qquad\qquad i=1,2,\ldots,n\ .
\end{aligned} \tag{5.64}$$

Only the members of the planetary system with non-zero masses were included in the initial value problem (5.64). This reduction is fully justified from the purely mathematical, not necessarily from the physical point of view – but the equations for celestial bodies of negligible mass may be treated with methods addressed in the previous section.

As indicated in eqns. (5.64) the osculating elements referring to the initial epoch t_0 were used to define the planets' initial state vectors.

There is only one dynamical parameter, namely the mass m_i, associated with each celestial body of the planetary system, which implies that "only" $n_p = 7n+1$ parameters define the particular solution of the initial value problem (5.64).

Generalizing the scheme set in section 5.2 the following notation for the parameters is used:

$$\begin{aligned}
\{p_1,p_2,\ldots,p_{7n+1}\} = \{\ & a_{10},e_{10},i_{10},\Omega_{10},\omega_{10},T_{10}, \\
& a_{20},e_{20},i_{20},\Omega_{20},\omega_{20},T_{20}, \\
& \ldots,\ldots,\ldots,\ldots,\ldots,\ldots, \\
& \ldots,\ldots,\ldots,\ldots,\ldots,\ldots, \\
& a_{n0},e_{n0},i_{n0},\Omega_{n0},\omega_{n0},T_{n0}, \\
& m_0,m_1,\ldots,m_n \qquad \}\ .
\end{aligned} \tag{5.65}$$

Let

$$p \in \{p_1,p_2,\ldots,p_{7n+1}\}$$

be one of the parameters of the system. Let us furthermore denote the partial derivative of the orbit $\boldsymbol{r}_i(t)$ of planet i (characterized by the planet's component matrix) w.r.t. the parameter p by

$$\boldsymbol{z}_i(t) \stackrel{\text{def}}{=} \left(\frac{\partial \boldsymbol{r}_i}{\partial p}\right)(t)\ , \quad i=1,2,\ldots,n\ . \tag{5.66}$$

For the entire planetary system we define the following column matrix (of dimension $d = 3\ n$) as the partial derivative of the solution vector of the entire system:

$$\boldsymbol{z}(t) \overset{\text{def}}{=} \begin{pmatrix} \boldsymbol{z}_1(t) \\ \boldsymbol{z}_2(t) \\ \dots \\ \dots \\ \boldsymbol{z}_n(t) \end{pmatrix} . \tag{5.67}$$

One column array of type (5.67) has to set up for each of the $7n+1$ parameters. The array (5.67) tells, what the impact of a small change in a particular parameter on the state vector of the entire planetary system is. Consider, e.g., the parameter $p \overset{\text{def}}{=} a_{10}$, the osculating semi-major axis of the innermost planet at t_0 in the system. The elements $10-12$ of the array (5.67) contain the partial derivatives of the components of the position vector of orbit $\boldsymbol{r}_4(t)$ of planet number 4 w.r.t. the semi-major axis of the first planet. The example shows that vector $\boldsymbol{z}(t)$ contains the complete (first order) information concerning the dependence of the entire system on the parameter considered.

The variational equations for the entire planetary system are obtained by taking the partial derivative of the primary equations (5.64) using the general procedure outlined in section 5.2. It is useful to introduce the following auxiliary 3×3 matrices to express the resulting variational equations in convenient form:

$$\mathbf{C}_{ij} \overset{\text{def}}{=} \frac{k^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3} \left[\mathbf{E} - \frac{3\,(\boldsymbol{r}_i - \boldsymbol{r}_j) \otimes (\boldsymbol{r}_i - \boldsymbol{r}_j)^T}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^2} \right] , \quad i,j = 1,2,\dots,n, \; j \neq i \tag{5.68}$$

and

$$\mathbf{C}_{i0} \overset{\text{def}}{=} \frac{k^2}{r_i^3} \left[\mathbf{E} - \frac{3\,(\boldsymbol{r}_i \otimes \boldsymbol{r}_i^T)}{r_i^2} \right] , \quad i = 1,2,\dots,n \; , \tag{5.69}$$

as well as

$$\mathbf{A}_{ij} \overset{\text{def}}{=} -\, m_j \big(-\mathbf{C}_{ij} + \mathbf{C}_{j0} \big) \; , \quad i,j = 1,2,\dots,n, \; j \neq i \tag{5.70}$$

and

$$\mathbf{A}_{ii} \overset{\text{def}}{=} -\,(m_0 + m_i)\,\mathbf{C}_{i0} \; - \sum_{j=1, j\neq i}^{n} m_j\,\mathbf{C}_{ij} \; , \quad i = 1,2,\dots,n \; . \tag{5.71}$$

The auxiliary matrices $\mathbf{A}_{ij}$, $i,j = 1,2,\dots,n$, (of dimension 3×3) are now arranged in one matrix of dimension $3n \times 3n$

$$\mathbf{A}_0 \overset{\text{def}}{=} \begin{pmatrix} \mathbf{A}_{11}, & \mathbf{A}_{12}, & \dots, & \mathbf{A}_{1n} \\ \mathbf{A}_{21}, & \mathbf{A}_{22}, & \dots, & \mathbf{A}_{2n} \\ \dots, & \dots, & \dots, & \dots \\ \dots, & \dots, & \dots, & \dots \\ \mathbf{A}_{n1}, & \mathbf{A}_{n2}, & \dots, & \mathbf{A}_{nn} \end{pmatrix} . \tag{5.72}$$

With these definitions the variational equations for the parameters associated with the initial osculating elements assume the standard form

$$\ddot{\boldsymbol{z}} = \mathbf{A}_0\, \boldsymbol{z}\ , \quad p \in \{p_1, p_2, \ldots, p_{6n}\}\ . \tag{5.73}$$

The initial values $\boldsymbol{z}(t_0),\ \dot{\boldsymbol{z}}(t_0)$ are obtained by taking the partial derivatives of the equations defining the initial conditions in the initial value problem (5.64). Only the formulas related to the two-body problem are required for that purpose.

Let us assume that p designates an osculating orbital element of planet number j. In this case, all the elements of $\boldsymbol{z}(t_0), \dot{\boldsymbol{z}}(t_0)$ *not* referring to this planet are zero. The elements of $\boldsymbol{z}_j(t_0), \dot{\boldsymbol{z}}_j(t_0)$ are calculated according to the procedure given in section 5.3.

If the parameter $p \in \{m_0, m_1, \ldots, m_n\}$ refers to one of the planetary masses or to the solar mass, the corresponding system of variational equations is inhomogeneous:

$$\ddot{\boldsymbol{z}} = \mathbf{A}_0\, \boldsymbol{z}\ +\ \boldsymbol{f}_p\ , \quad p \in \{m_0, m_1, \ldots, m_n\}\ . \tag{5.74}$$

For $i > 0$ the parameter $p \overset{\text{def}}{=} m_i$ is one of the planetary masses and the vector $\boldsymbol{f}_p$ assumes the form

$$\boldsymbol{f}_p = \begin{pmatrix} -\,\delta_1^i\, k^2\, \dfrac{\boldsymbol{r}_1}{r_1^3}\ -\ k^2 \displaystyle\sum_{j=2}^{n} \delta_j^i \left[\dfrac{\boldsymbol{r}_1 - \boldsymbol{r}_j}{|\boldsymbol{r}_1 - \boldsymbol{r}_j|^3} + \dfrac{\boldsymbol{r}_j}{r_j^3} \right] \\ -\,\delta_2^i\, k^2\, \dfrac{\boldsymbol{r}_2}{r_2^3}\ -\ k^2 \displaystyle\sum_{j=1, j\neq 2}^{n} \delta_j^i \left[\dfrac{\boldsymbol{r}_2 - \boldsymbol{r}_j}{|\boldsymbol{r}_2 - \boldsymbol{r}_j|^3} + \dfrac{\boldsymbol{r}_j}{r_j^3} \right] \\ \cdots \\ \cdots \\ -\,\delta_n^i\, k^2\, \dfrac{\boldsymbol{r}_n}{r_n^3}\ -\ k^2 \displaystyle\sum_{j=1}^{n-1} \delta_j^i \left[\dfrac{\boldsymbol{r}_n - \boldsymbol{r}_j}{|\boldsymbol{r}_n - \boldsymbol{r}_j|^3} + \dfrac{\boldsymbol{r}_j}{r_j^3} \right] \end{pmatrix}, \tag{5.75}$$

$$p \in \{m_1, m_2, \ldots, m_n\}\ .$$

δ_i^k is the *Kronecker-symbol* (named after Leopold Kronecker (1823–1891)), assuming the values $\delta_i^k = 0$ for $i \neq k$ and $\delta_i^k = 1$ for $i = k$. Observe, that each element of vector $\boldsymbol{f}_p$ consists only of one term (either stemming from the main term or from one of the terms in the sum of the perturbing accelerations).

For $p = m_0$ (=solar mass), $\boldsymbol{f}_p$ assumes the form:

$$\boldsymbol{f}_p = \begin{pmatrix} -\,k^2\, \dfrac{\boldsymbol{r}_1}{r_1^3} \\ -\,k^2\, \dfrac{\boldsymbol{r}_2}{r_2^3} \\ \cdots \\ \cdots \\ -\,k^2\, \dfrac{\boldsymbol{r}_n}{r_n^3} \end{pmatrix}\ . \tag{5.76}$$

For dynamical parameters $p \in \{m_0, m_1, \ldots, m_n\}$ the variational equations (5.74) are inhomogeneous, but the corresponding initial conditions are homogeneous (all zero): $\boldsymbol{z}(t_0) = \boldsymbol{0}$ and $\dot{\boldsymbol{z}}(t_0) = \boldsymbol{0}$.

Formally, it is rather simple to implement the solution of all variational equations into a computer program like PLASYS (see Chapter II- 10 of Part III). The computational effort and the data handling aspect should not be underestimated, however: Instead of solving one differential equation system of order 2 and dimension $3n$ (n being the number of planets included in the integration), we would have to solve $+7n$ such systems, where all systems, with the exception of the primary system, are linear. When analyzing the outer planetary system – without Pluto – we would therefore roughly increase the processing time requirements by a factor of 30, and the storage requirements would increase by a similar factor.

The benefits of including the system of variational equations are, on the other hand, considerable: we do not only obtain information related to the development of a sample planetary system, but the complete information concerning its (first order) stability within the time interval of the integration. Exactly as in the case of a minor planet (see Chapter II- 4.3) the variational equations might serve to look for a chaotic behavior in the development of a planetary system. It would also be extremely interesting to investigate the stability of the system with respect to the planetary masses.

Instead of trying to integrate the equations of motion of the planetary system over longer and longer time intervals, it would perhaps make more sense to perform integrations over moderately long intervals, let us say up to about 100 million years, but to include the variational equations into the integration process. Methods to implement such schemes without increasing the CPU requirements dramatically will be discussed in the next section.

5.6 Efficient Solution of the Variational Equations

The general structure of the variational equations were developed in section 5.2, the structure of variational equations of Celestial Mechanics were then discussed in sections 5.4 and 5.5. The distinction was made between the motion of an individual body in a given force field (e.g., minor planet of negligible mass or an artificial Earth satellite) and the solution of the N-body problem governed by an initial value problem of type (5.64).

In section 5.2 we saw (see eqns. (5.21)) that there are powerful methods to solve the variational equations associated with dynamical parameters. So far, the issue of solving the variational equations associated with initial values (e.g., the osculating elements at t_0) was not addressed.

The development of efficient techniques to solve the variational equations with this parameter type is the purpose of the current section. The specific structure of the equations of motion of Celestial Mechanics is exploited for this purpose. The methods outlined below are therefore not literally transferrable to dynamical problems outside Celestial Mechanics.

5.6.1 Trajectories of Individual Bodies

In the most general case (disregarding dynamical parameters) the variational equations

$$\ddot{\boldsymbol{z}} = \big(\mathbf{A}_{00}(\boldsymbol{r}) + \varDelta\mathbf{A}_0\big)\,\boldsymbol{z} + \mathbf{A}_1\,\dot{\boldsymbol{z}} \tag{5.77}$$

have to be solved, where the matrix of the two-body problem $\mathbf{A}_{00}(t)$ is given by eqn. (5.23), and where the elements of the matrices $\varDelta\mathbf{A}_0$ and $\mathbf{A}_1$ are small quantities when compared to the elements of matrix $\mathbf{A}_{00}$. We assume that the partial derivative $\boldsymbol{z}(t)$ refers to one of the osculating elements at t_0:

$$p \in \{a_0, e_0, i_0, \varOmega_0, \omega_0, T_{00}\}\ .$$

The initial values at time t_0 are:

$$\boldsymbol{z}(t_0) = \frac{\partial \boldsymbol{r}_0}{\partial p} \quad \text{and} \quad \dot{\boldsymbol{z}}(t_0) = \frac{\partial \dot{\boldsymbol{r}}_0}{\partial p}\ . \tag{5.78}$$

Except for simple special cases, the above initial value problem has to be solved by numerical methods. The prominent exception is the two-body problem with $\varDelta\mathbf{A}_0 = \mathbf{A}_1 = \mathbf{0}$, which was discussed in section 5.3.

Let us therefore define an auxiliary initial value problem, which differs from problem (5.77), (5.78) only by the primary and variational differential equations, which are those of the two-body problem. Let us furthermore assume that $\boldsymbol{z}_0(t)$ solves this auxiliary problem:

$$\begin{aligned} \ddot{\boldsymbol{z}}_0 \quad &= \mathbf{A}_{00}(\boldsymbol{r}_0)\,\boldsymbol{z}_0 \\ \boldsymbol{z}_0(t_0) &= \frac{\partial \boldsymbol{r}_0}{\partial p} \quad \text{and} \quad \dot{\boldsymbol{z}}_0(t_0) = \frac{\partial \dot{\boldsymbol{r}}_0}{\partial p}\ . \end{aligned} \tag{5.79}$$

$\boldsymbol{z}_0(t)$ thus is the solution of the variational equations associated with the two-body problem obeying the same initial conditions at t_0 as the function $\boldsymbol{z}(t)$.

Introducing the notation

$$\varDelta\boldsymbol{z}(t) \overset{\text{def}}{=} \boldsymbol{z}(t) - \boldsymbol{z}_0(t) \tag{5.80}$$

for the difference between the two partial derivatives $\boldsymbol{z}(t)$ and $\boldsymbol{z}_0(t)$ one may easily establish a differential equation system for this difference

$$\varDelta\ddot{\boldsymbol{z}} = \mathbf{A}_{00}(\boldsymbol{r}_0)\,\varDelta\boldsymbol{z} \,+\, \big(\varDelta\mathbf{A}_0 + \delta\mathbf{A}_{00}\big)\,\boldsymbol{z} \,+\, \mathbf{A}_1\,\dot{\boldsymbol{z}}\ , \tag{5.81}$$

where $\delta\mathbf{A}_{00} \overset{\text{def}}{=} \mathbf{A}_{00}(\boldsymbol{r}) - \mathbf{A}_{00}(\boldsymbol{r}_0)$ is the difference of matrices $\mathbf{A}_{00}$, as computed with formula (5.23), once with actual $\boldsymbol{r}(t)$, once with the two-body approximation $\boldsymbol{r}_0(t)$.

Because

$$\Delta\boldsymbol{z}(t_0) = \mathbf{0} \quad \text{and} \quad \Delta\dot{\boldsymbol{z}}(t_0) = \mathbf{0} \tag{5.82}$$

and because the matrices $\Delta\mathbf{A}_0 + \delta\mathbf{A}_0$ and $\mathbf{A}_1$ are small when compared to the matrix $\mathbf{A}_{00}$, $\Delta\boldsymbol{z}(t)$ is a small quantity in the vicinity of the initial epoch t_0, as well.

This in turn implies that the second and third term in eqn. (5.81) are small when compared to the first term. Equation (5.81) therefore may be solved by the following iteration process:

$$\Delta\ddot{\boldsymbol{z}}^{[I+1]} = \mathbf{A}_{00}\,\Delta\boldsymbol{z}^{[I+1]} + \left(\delta\mathbf{A}_{00} + \Delta\mathbf{A}_0\right)\boldsymbol{z}^{[I]} + \mathbf{A}_1\,\dot{\boldsymbol{z}}^{[I]}\ , \quad I = 1, 2, \ldots\ . \tag{5.83}$$

The process is initialized for $I = 1$ by

$$\boldsymbol{z}^{[I]}(t) = \boldsymbol{z}_0(t)\ . \tag{5.84}$$

Using eqn. (5.84) in eqn. (5.83) (which corresponds to the first approximation step) actually is a first order approximation in the spirit of perturbation theory (see Chapter 6).

We have thus replaced the solution of the system of variational equations (5.77) by an iterative solution of the system of equations (5.83), which has the advantage that the solution of the corresponding homogeneous system (as it is the system of variational equations corresponding to the two-body problem) is known: Its complete solution is given by the functions (5.21), where the inhomogeneous part has to be adapted to the structure of eqns. (5.83). We have therefore shown that the solution of the variational equations of the perturbed motion may be reduced to the calculation of definite integrals, a process which is orders of magnitude more efficient than the solution of differential equations.

Usually, the iterative solution process (5.83) may be terminated after the first step, which corresponds to a solution in the tradition of the perturbation theory of the first order.

From the implementation point of view it is simpler to solve the variational equations simultaneously with the primary equations without performing any transformations (this is why this method of solving the variational equations was implemented in program PLASYS), but from the economical point of view such a procedure cannot be recommended. The procedure described above was followed in program SATORB, when used in the orbit determination mode (see Chapter II- 7 of Part III).

5.6.2 The *N*-Body Problem

A closer inspection of the homogeneous system of variational equations (5.73) associated with the initial values of the primary equations (5.64) of a planetary system reveals, that an approximative solution based essentially on the same principles as those developed for the trajectories of individual bodies in a pre-determined field is feasible.

According to eqn. (5.73) the variational equations referring to an initial osculating element of one of the bodies may be written as

$$\ddot{\boldsymbol{z}} = \mathbf{A}_0 \, \boldsymbol{z} \ . \tag{5.85}$$

The structure (5.72) of matrix $\mathbf{A}_0$ shows that all but the diagonal matrix elements $\mathbf{A}_{ii}$ are small quantities proportional to one mass m_j or to a linear combination of terms, each of which is proportional to a planetary mass. The matrix may be decomposed as follows:

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{A}_{001}, & \mathbf{0}, & \ldots, & \mathbf{0} \\ \mathbf{0}, & \mathbf{A}_{002}, & \ldots, & \mathbf{0} \\ \ldots, & \ldots, & \ldots, & \ldots \\ \ldots, & \ldots, & \ldots, & \ldots \\ \mathbf{0}, & \mathbf{0}, & \ldots, & \mathbf{A}_{00n} \end{pmatrix} + \begin{pmatrix} \delta\mathbf{A}_{11}, & \mathbf{A}_{12}, & \ldots, & \mathbf{A}_{1n} \\ \mathbf{A}_{21}, & \delta\mathbf{A}_{22}, & \ldots, & \mathbf{A}_{2n} \\ \ldots, & \ldots, & \ldots, & \ldots \\ \ldots, & \ldots, & \ldots, & \ldots \\ \mathbf{A}_{n1}, & \mathbf{A}_{n2}, & \ldots, & \delta\mathbf{A}_{nn} \end{pmatrix} , \tag{5.86}$$

where

$$\mathbf{A}_{00i} \overset{\text{def}}{=} -\frac{k^2\,(m_0+m_i)}{r_i^3}\left[\mathbf{E} - \frac{3\,(\boldsymbol{r}_i \otimes \boldsymbol{r}_i^T)}{r_i^2}\right] , \quad i = 1,2,\ldots,n \tag{5.87}$$

and

$$\delta\mathbf{A}_{ii} \overset{\text{def}}{=} -\sum_{j=1, j\neq i}^{n} m_j \, \mathbf{C}_{ij} \ , \quad i = 1,2,\ldots,n \ . \tag{5.88}$$

The matrices $\mathbf{A}_{00i}$ are the matrix of the variational equations of the two-body problem with the total mass $m_0 + m_i$ of the system (and the associated gravity constant $k^2\,(m_0+m_i)$). All the matrix elements of the above matrix $\mathbf{A}_0$, with the exception of the matrices $\mathbf{A}_{00i}$, are small quantities of the first order in the planetary masses m_j , $j = 1,2,\ldots,n$.

The differential equations referring to one of the planets, say planet j, are of the same structure as those referring to an individual trajectory (established in the previous paragraph), which is why these equations may be solved with the pattern established in the previous paragraph. For that purpose we introduce $\boldsymbol{z}_{0i}$ as the solution of variational equations associated with the two-body problem of planet i

$$\boldsymbol{z}_{0i} = \mathbf{A}_{00i} \, \boldsymbol{z}_{0i} \ . \tag{5.89}$$

Designating by $\Delta \boldsymbol{z}(t) = \boldsymbol{z}_i(t) - \boldsymbol{z}_{0i}(t)$ the difference between the actual and the two-body version of the variational equations (obeying the same initial conditions) for planet i, the following differential equation system is obtained:

$$\Delta \ddot{\boldsymbol{z}}_i^{[I+1]} = \mathbf{A}_{00i}\, \Delta \boldsymbol{z}_i^{[I+1]} + \sum_{i=1,j\neq i}^{n} \mathbf{A}_{ij}\, \boldsymbol{z}_j^{[I]} + \left(\delta \mathbf{A}_{ii} + \delta \mathbf{A}_{00i}\right) \boldsymbol{z}_i^{[I]} \ , \tag{5.90}$$

where $\delta \mathbf{A}_{00i} \overset{\text{def}}{=} \mathbf{A}_{00}(\boldsymbol{r}_i) - \mathbf{A}_{00}(\boldsymbol{r}_{0i})$ is computed with formula (5.23) using the correct and the two-body solution of the initial value problem (5.85, 5.86) for planet j.

For each planet the solution of the above linear, inhomogeneous system of equations is performed using the routine procedure set up for individual trajectories: A complete solution of the homogeneous system is produced first (this solution is obtained in closed form as the solution of the variational equation accompanying the two-body problem). The solution of the inhomogeneous equation is obtained afterwards by the method of variation of constants.

5.7 Variational Equations and Error Propagation

The numerical solution of an initial value problem of type (5.1) differs from its true solution due to small errors of different kind, introduced at discrete epochs t_k , $k = 0, 1, 2, \ldots$, into true state vector (see Chapter 7). The difference *numerically integrated – true solution* may be written with the help of the complete system of solutions of the homogeneous system (5.12) of variational equations associated with the initial value problem (5.1). In this section we develop the explicit form of this representation.

Let us assume that at epoch $t = t_k$ the errors $\boldsymbol{\varepsilon}_k^i$ are introduced into the state vector $\boldsymbol{y}^{(i)}(t)$, $i = 0, 1, \ldots, n-1$. The error $\Delta \boldsymbol{z}_k(t)$ of the true state vector at a time $t \geq t_k$ due to the errors $\boldsymbol{\varepsilon}_k^i$ introduced at t_k is defined as the solution of the initial value problem

$$\begin{aligned} \Delta \boldsymbol{z}_k^{(n)} \quad &= \sum_{i=0}^{n-1} \mathbf{A}_i(t)\, \Delta \boldsymbol{z}_k^{(i)} \\ \Delta \boldsymbol{z}_k^{(i)}(t_k) &= \boldsymbol{\varepsilon}_k^i \ , \quad i = 0, 1, \ldots, n-1 \ , \end{aligned} \tag{5.91}$$

where the matrices $\mathbf{A}_i(t)$ are defined by eqn. (5.7). Equations (5.91) represent nothing but a particular solution of the homogeneous part of the variational equations (5.6) associated with the initial value problem (5.1).

The solution of the above initial value problem may be written as a linear combination of type (5.16) (with constant coefficients, however) of the complete system of solutions of the homogeneous system of variational equations associated with problem (5.1):

$$\Delta \boldsymbol{z}_k^{(i)}(t) = \mathbf{Z}^{(i)}(t)\,\boldsymbol{\alpha}_k\ , \quad i = 0, 1, \ldots, n-1\ , \tag{5.92}$$

where $\mathbf{Z}(t)$ is the rectangular matrix with nd columns and d rows, in which column j contains the elements of the solution $\boldsymbol{z}_j(t)$ with index j of the homogeneous system (5.12) (see eqn. (5.16)).

The nd (constant) coefficients in array $\boldsymbol{\alpha}_k$ are determined by the request that the linear combination (5.92) must satisfy the initial conditions in eqns. (5.91):

$$\Delta \boldsymbol{z}_k^{(i)}(t_k) = \mathbf{Z}^{(i)}(t_k)\,\boldsymbol{\alpha}_k = \boldsymbol{\varepsilon}_k^i\,, \quad i = 0, 1, \ldots, n-1\ . \tag{5.93}$$

Using the notations

$$\begin{aligned} \Delta\tilde{\mathbf{Z}}_k^T(t) &\overset{\text{def}}{=} \Delta\left(\boldsymbol{z}_k^T(t), \Delta\dot{\boldsymbol{z}}_k^T(t), \ldots, \left(\Delta\boldsymbol{z}_k^{(n)}(t)\right)^T\right) \\ \tilde{\boldsymbol{\varepsilon}}_k^T &= \left(\left(\boldsymbol{\varepsilon}_k^0\right)^T, \left(\boldsymbol{\varepsilon}_k^1\right)^T, \ldots, \left(\boldsymbol{\varepsilon}_k^{n-1}\right)^T\right)\ , \end{aligned} \tag{5.94}$$

the above condition equations may be given the compact matrix form

$$\tilde{\mathbf{Z}}(t_k)\,\boldsymbol{\alpha}_k = \tilde{\boldsymbol{\varepsilon}}_k\ , \tag{5.95}$$

allowing it to determine the coefficient matrix as:

$$\boldsymbol{\alpha}_k = \tilde{\mathbf{Z}}^{-1}(t_k)\,\tilde{\boldsymbol{\varepsilon}}_k\ . \tag{5.96}$$

This result allows it in turn to write the solution of the initial value problem (5.91) as:

$$\Delta\tilde{\mathbf{Z}}_k(t) = \tilde{\mathbf{Z}}(t)\ \tilde{\mathbf{Z}}^{-1}(t_k)\ \tilde{\boldsymbol{\varepsilon}}_k\ . \tag{5.97}$$

Equation (5.97) describes the development of the errors introduced at one particular epoch t_k (the grid points of the numerical solution of an initial value problem) as a function of time. The accumulated error at t_N due to *all* errors introduced at all the epochs t_k , $k = 1, 2, \ldots, N$, may then simply be calculated as the superposition of all errors (5.97) evaluated at t_N:

$$\Delta\tilde{\mathbf{Z}}(t_N) \overset{\text{def}}{=} \sum_{i=0}^{N} \Delta\tilde{\mathbf{Z}}_k(t_N) = \tilde{\mathbf{Z}}(t_N) \sum_{k=0}^{N} \tilde{\mathbf{Z}}^{-1}(t_k)\,\tilde{\boldsymbol{\varepsilon}}_k\ . \tag{5.98}$$

Equation (5.98) may be called the *fundamental law of error propagation.* It represents the accumulated integration error as a linear combination of the elementary errors introduced into the state vector at epochs t_k. The result will be used to describe the accumulation of rounding and approximation errors in Chapter 7.

6. Theory of Perturbations

6.1 Motivation and Classification

The expression *perturbed motion* implies that there is an *unperturbed motion.* In Celestial Mechanics the unperturbed motion is the orbital motion of two spherically symmetric bodies represented by the equations of motion (4.1), the solution of which is known in terms of simple analytical functions (see section 4.1). The constant μ is the product of the constant of gravitation and the sum of the masses of the two bodies considered. The numerical value of μ thus depends on the concrete problem and on the system of units chosen.

The *perturbed motion* of a celestial body is defined as the solution of an initial value problem of the following type:

$$\ddot{\boldsymbol{r}} = -\,\mu\,\frac{\boldsymbol{r}}{r^3} \,+\, \delta\boldsymbol{f}(t,\boldsymbol{r},\dot{\boldsymbol{r}}) \ , \tag{6.1}$$

$$\boldsymbol{r}(t_0) = \boldsymbol{r}_0 \quad \text{and} \quad \dot{\boldsymbol{r}}(t_0) = \boldsymbol{v}_0 \ . \tag{6.2}$$

For a system of point masses there is one such equation for each of the bodies, except for the central body to which the position vectors refer.

The term $-\,\mu\,\frac{\boldsymbol{r}}{r^3}$ in eqn. (6.1) is called the *two-body term,* $\delta\boldsymbol{f}$ the *perturbation term.* The terminology makes sense if the perturbation term is considerably smaller than the two-body term, i.e., if

$$\mid \delta\boldsymbol{f} \mid \ll \left| -\,\mu\,\frac{\boldsymbol{r}}{r^3} \right| \ . \tag{6.3}$$

Three different kinds of perturbation equations were introduced in Chapter 3. One easily verifies that condition (6.3) is met for the planetary N-body problem represented by eqns. (3.18), for the motion of a body with negligible mass in the planetary system (represented by eqn. (3.21)), for the geocentric motion (3.118) of Moon and Sun in the generalized three-body problem Earth-Sun-Moon, and for the motion of an artificial Earth satellite represented by eqn. (3.143). Observe that the general relativistic equations of motion (3.186) or their "light" version (3.190) also have the same structure as eqns. (6.1) and that condition (6.3) is easily met, allowing us to consider the relativistic two-body problem (see section 4.4) as a perturbed classical two-body problem.

The differential equation system (6.1) is called the *system of perturbation equations* or simply the *perturbation equations.* Every method solving the initial value problem (6.1, 6.2) is a called a *perturbation method.*

In Celestial Mechanics one usually makes the distinction between

- *General Perturbation Methods*, seeking the solution in terms of series of elementary integrable functions, and
- *Special Perturbation Methods*, seeking at some stage the solution by the methods of numerical integration.

For general perturbation methods it is mandatory *not* to use the original equations of motion (6.1) in rectangular coordinates, *but* to derive differential equations for the osculating orbital elements (see section 4.3) or for functions thereof. This procedure promises to make the best possible use of the (analytically known) solution of the two-body problem (4.1), because the osculating elements are so-called first integrals of the two-body motion.

Both, general and special perturbation methods, provide approximate solutions of the equations of motion (not regarding the few special cases which could be solved in closed form). In the former case the approximation is due to the fact that the series developments have to be terminated at some point and that sometimes the convergence of the series is not well established, in the latter case it is due to the accumulation of rounding and approximation errors to be discussed in Chapter 7.

Special perturbation methods may be applied directly to the initial value problem (6.1, 6.2) or to the transformed equations for the osculating elements. Solution algorithms are discussed in Chapter 7.

In this Chapter the focus is on transformations of the initial value problem (6.1, 6.2) with the goal to make optimum use of the analytical solution of the two-body problem (4.1). In section 6.2 a differential equation is developed for the difference vector of a perturbed and the associated unperturbed motion (obeying eqns. (6.1) and (4.1), respectively, both meeting the same initial conditions (6.2)). The analytical developments necessary for this purpose are rather moderate, the importance is considerable in practice. In section 6.3 we outline the method to derive the differential equations for the osculating elements starting from the original equations of motion (6.1). The perturbation term $\delta \boldsymbol{f}$ may be rather arbitrary. The resulting equations usually are referred to as the *Gaussian perturbation equations.*

In section 6.4 the perturbation equations are derived under the assumption that the perturbation term may be written as the gradient of a scalar force function. The resulting equations are called *Langrange's planetary equations.* The Gaussian and Lagrangian perturbation equations are derived directly from the Newton-Euler equations of type (6.1) without making use of the results of analytical mechanics. First- and higher-order perturbation methods are discussed in section 6.5. When applying general perturbation methods, the scalar perturbation function has to be transformed into a form allowing for an analytical integration. This task, which may be frustratingly complicated is briefly addressed in section 6.6. The Gaussian and Lagrangian versions of the perturbation equations are set up and solved for the so-called osculating orbital elements of the celestial bodies considered. These elements are first integrals (integration constants) of the two-body problem. Having solved that problem it is possible to set up differential equations for functions of these orbital elements. It turns out that the equation for the mean anomaly $\sigma(t) \overset{\text{def}}{=} n(t)(t-T_0)$ obeys a particularly simple equation, which is, as a matter of fact, preferable to the equations for the time T_0 of pericenter passage or for the mean anomaly $\sigma_0 \overset{\text{def}}{=} \sigma(t_0)$ at the initial epoch t_0. The equations for σ are developed for the Gaussian and Lagrangian version in section 6.7.

6.2 Encke-Type Equations of Motion

A simple method to solve the initial value problem (6.1, 6.2) making intelligent use of the solution of the two-body problem (4.1) is attributed to the German astronomer Johann Franz Encke (1791–1865). It is based on a differential equation for the difference vector $\Delta\boldsymbol{r}(t) \overset{\text{def}}{=} \boldsymbol{r}(t) - \boldsymbol{r}_0(t)$, where $\boldsymbol{r}(t)$ is the solution of the perturbed motion (6.1), $\boldsymbol{r}_0(t)$ the solution of the corresponding two-body motion (4.1), and where both solution vectors assume the same initial values (6.2). The method is not only well suited to describe the motion of an individual particle in a given force field (e.g., of a minor planet or of an artificial Earth satellite) it may also be adapted to the integration of the entire planetary system. Encke's method is well established in astronomy. It was, e.g., used to integrate the planetary equations of motion in the LONGSTOP project [95].

With equations (6.1), (4.1) and the common initial values (6.2) the initial value problem for the difference vector $\Delta\boldsymbol{r}(t) \overset{\text{def}}{=} \boldsymbol{r}(t) - \boldsymbol{r}_0(t)$ is easily set up:

$$\begin{aligned} \Delta\ddot{\boldsymbol{r}}_0 &= -\mu\left\{\frac{\boldsymbol{r}_0+\Delta\boldsymbol{r}}{|\boldsymbol{r}_0+\Delta\boldsymbol{r}|^3} - \frac{\boldsymbol{r}_0}{r_0^3}\right\} + \delta\boldsymbol{f}\big(t, \boldsymbol{r}_0+\Delta\boldsymbol{r}, \dot{\boldsymbol{r}}_0+\Delta\dot{\boldsymbol{r}}\big) \\ \Delta\boldsymbol{r}_0(t_0) &= \boldsymbol{0} \\ \Delta\dot{\boldsymbol{r}}_0(t_0) &= \boldsymbol{0}\ , \end{aligned} \tag{6.4}$$

where vector $\boldsymbol{r}_0(t)$ and its derivative on the right-hand side of the differential equation are the known solutions of the two-body problem.

The term in brackets $\{\ldots\}$ in the above differential equation system is a small quantity in the vicinity of the initial epoch t_0 (due to the initial values of the difference vector $\Delta \boldsymbol{r}(t)$), the right-hand side of the equation is a small quantity, because the perturbation term $\delta \boldsymbol{f}$ is small, as well. In the formulation (6.4) the term is calculated as the difference of two "large" quantities (compared to the result). It is therefore advisable, to look for an alternative representation of the term $\{\ldots\}$.

Using the notation of Brouwer and Clemence [27] we obtain:

$$-\mu \left\{ \frac{\boldsymbol{r}_0 + \Delta \boldsymbol{r}}{|\boldsymbol{r}_0 + \Delta \boldsymbol{r}|^3} - \frac{\boldsymbol{r}_0}{r_0^3} \right\} = -\frac{\mu}{r_0^3} \{ \Delta \boldsymbol{r} - f q \, (\boldsymbol{r}_0 + \Delta \boldsymbol{r}) \} \ , \tag{6.5}$$

where

$$q = \frac{1}{r_0^2} \left(\boldsymbol{r}_0 + \frac{1}{2} \Delta \boldsymbol{r} \right) \cdot \Delta \boldsymbol{r} \tag{6.6}$$

and

$$f = \frac{1 - (1 + 2\,q)^{-3/2}}{q} \ . \tag{6.7}$$

The initial value problem to be solved when using Encke's formulation is obtained by replacing the brackets $\{\ldots\}$ in the differential equations (6.4) by the expression (6.5):

$$\begin{aligned} \Delta \ddot{\boldsymbol{r}}_0 &= -\frac{\mu}{r_0^3} \{ \Delta \boldsymbol{r} - f q \, (\boldsymbol{r}_0 + \Delta \boldsymbol{r}) \} + \delta \boldsymbol{f}\big(t, \boldsymbol{r}_0 + \Delta \boldsymbol{r}, \dot{\boldsymbol{r}}_0 + \Delta \dot{\boldsymbol{r}}\big) \\ \Delta \boldsymbol{r}_0(t_0) &= \boldsymbol{0} \\ \Delta \dot{\boldsymbol{r}}_0(t_0) &= \boldsymbol{0} \ . \end{aligned} \tag{6.8}$$

Due to the factors q and f (which are non-linear functions of $\Delta \boldsymbol{r}$) the differential equation in the initial value problem (6.8) is non-linear and of second order. It must be solved by numerical integration (see Chapter 7), i.e., by special perturbation methods.

When compared to the direct integration of the initial value problem (6.1, 6.2), the solution of the initial value problem (6.8) is perhaps twice as efficient. A gain of this kind matters only, if the problem has to be integrated over long time spans. This is the case when integrating the planetary system over millions of years as it was, e.g., done in the LONGSTOP project [95]. The decision to use Encke's method was undoubtedly beneficial to the project.

One should on the other hand keep in mind, that the use of eqns. (6.8) is only advantageous in the vicinity of the initial epoch t_0. After a few revolutions it is no longer justified to consider the vector $\Delta \boldsymbol{r}$ as a small quantity and Encke's method would loose its efficiency. It is therefore necessary to "re-initialize" Encke's method from time to time by introducing new initial epochs $\tilde{t}_{0i}$ and by defining new initial value problems by replacing t_0 by $\tilde{t}_{0i}$

in eqns. (6.8), a process which involves (among other) the computation of new osculating elements. From the point of view of processing time this does not matter (because such re-initialization events are only required after a few hundred integration steps) but the programming logics required should not be underestimated. If processing time is not a critical issue there is no point in using Encke's method.

There is an alternative method to solve the initial value problem (6.4) promising to be much more efficient. The method reduces the problem from one of solving differential equations to one of calculating definite integrals (i.e., to quadrature). The gain resides in the fact that there are much more efficient methods for numerical quadrature than for the solution of ordinary differential equations (see Chapter 7).

The method transforms the equations (6.4) as follows: The term in brackets $\{\ldots\}$ is correctly linearized in the small quantity $\Delta\boldsymbol{r}$; the difference between the nonlinear and the linearized function $\{\ldots\}$ is designated by $\delta\boldsymbol{f}_0$:

$$-\mu\left\{\frac{\boldsymbol{r}_0+\Delta\boldsymbol{r}}{|\boldsymbol{r}_0+\Delta\boldsymbol{r}|^3}-\frac{\boldsymbol{r}_0}{r_0^3}\right\}=\mathbf{A}_{00}\,\Delta\boldsymbol{r}\;+\;\delta\boldsymbol{f}_0(\boldsymbol{r}_0,\Delta\boldsymbol{r})\;, \tag{6.9}$$

where the square 3 × 3-matrix

$$\mathbf{A}_{00}=-\frac{\mu}{r_0^3}\left\{\boldsymbol{E}-\frac{3}{r_0^2}\,\boldsymbol{r}_0\otimes\boldsymbol{r}_0^T\right\} \tag{6.10}$$

is identical with the matrix of the variational equations associated with the two-body problem as defined in eqns. (5.22), (5.23) in Chapter 5.

Equation (6.9) defines the function $\delta\boldsymbol{f}_0$. By virtue of this definition, $\delta\boldsymbol{f}_0$ is a small quantity of second order in vector $\Delta\boldsymbol{r}$. By introducing eqn. (6.9) into the differential equation (6.4) we obtain the following initial value problem:

$$\begin{aligned}\Delta\ddot{\boldsymbol{r}}_0\quad &=\mathbf{A}_{00}\,\Delta\boldsymbol{r}\;+\;\delta\boldsymbol{f}_0(\boldsymbol{r}_0,\Delta\boldsymbol{r})\;+\;\delta\boldsymbol{f}\big(t,\boldsymbol{r}_0+\Delta\boldsymbol{r},\dot{\boldsymbol{r}}_0+\Delta\dot{\boldsymbol{r}}\big)\\ \Delta\boldsymbol{r}_0(t_0)&=\boldsymbol{0}\\ \Delta\dot{\boldsymbol{r}}_0(t_0)&=\boldsymbol{0}\;.\end{aligned} \tag{6.11}$$

It is important to note that eqns. (6.11) are algebraically identical with Encke's original eqns. (6.4), i.e., the solution of the initial value problem is mathematically equivalent to the solution of the initial value problems (6.1, 6.2) and (6.8).

As already mentioned, the elements of $\delta\boldsymbol{f}_0$ are small quantities of the second order in $\Delta\boldsymbol{r}$. The perturbation term $\delta\boldsymbol{f}$, on the other hand, is a small quantity of the first-order. If we would use the approximation $\Delta\boldsymbol{r}\overset{\text{def}}{=}\boldsymbol{0}$ in the perturbation term $\delta\boldsymbol{f}$ we would therefore only neglect terms of the second order in small quantities.

This particular structure of the initial value problem (6.11) allows it to set up the following iterative solution process:

$$\begin{aligned}\varDelta\ddot{\boldsymbol{r}}_0^{[I+1]} &= \mathbf{A}_{00}\,\varDelta\boldsymbol{r}^{[I+1]} + \delta\boldsymbol{f}_0\Big(\boldsymbol{r}_0, \varDelta\boldsymbol{r}^{[I]}\Big) + \delta\boldsymbol{f}\Big(t, \boldsymbol{r}_0 + \varDelta\boldsymbol{r}^{[I]}, \dot{\boldsymbol{r}}_0 + \varDelta\dot{\boldsymbol{r}}^{[I]}\Big) \\ &\overset{\text{def}}{=} \mathbf{A}_{00}\,\varDelta\boldsymbol{r}^{[I+1]} + \delta\tilde{\boldsymbol{f}}^{[I]}\ , \quad I = 0, 1, \ldots \\ \varDelta\boldsymbol{r}_0(t_0) &= \mathbf{0} \\ \varDelta\dot{\boldsymbol{r}}_0(t_0) &= \mathbf{0}\ ,\end{aligned} \tag{6.12}$$

where

$$\varDelta\boldsymbol{r}^{[I]} = \mathbf{0} \quad \text{for } I = 0\ . \tag{6.13}$$

Note, that the differential equation system in eqns. (6.12) is a linear system of second order. The homogeneous part is the same as that of the two-body problem (see eqns. (5.22), Chapter 5.2), the inhomogeneous part $\delta\tilde{\boldsymbol{f}}$ consists of the two contributions $\delta\boldsymbol{f}_0$ and $\delta\boldsymbol{f}$.

The advantage of eqns. (6.12) over Encke's original equations (6.8) resides in the facts that

- the differential equation system (6.12) is a *linear*, inhomogeneous system, and that
- a complete system of the homogeneous part in this linear system is known in closed form. The solution was given explicitly in Chapter 5.

These two facts allow it, eventually, to reduce the solution of the inhomogeneous system to numerical quadrature: The six partial derivatives

$$\boldsymbol{z}_1(t) \overset{\text{def}}{=} \left(\frac{\partial\boldsymbol{r}_0}{\partial a}\right)(t)\,,\ \boldsymbol{z}_2(t) \overset{\text{def}}{=} \left(\frac{\partial\boldsymbol{r}_0}{\partial e}\right)(t)\,,\ \ldots,\ \boldsymbol{z}_6(t) \overset{\text{def}}{=} \left(\frac{\partial\boldsymbol{r}_0}{\partial T_0}\right)(t)$$

w.r.t. the osculating orbital elements a, e, i, $\varOmega$, ω, and T_0 at epoch t_0 form a complete system of solutions of the homogeneous equations associated with eqns. (6.12) (which are known in closed form). The solution of the inhomogeneous solution is then obtained by the method of variation of constants, as outlined in section 5.1.

The solution of the inhomogeneous equation (6.12) may be written as a linear combination of the homogeneous solutions, where the coefficients of the linear combination are time-dependent:

$$\varDelta\ddot{\boldsymbol{r}}_0^{[I+1]} = \mathbf{Z}\,\boldsymbol{\alpha}^{[I+1]}(t)\ , \tag{6.14}$$

where

$$\mathbf{Z}(t) = \big(\boldsymbol{z}_1(t), \boldsymbol{z}_2(t), \ldots, \boldsymbol{z}_6(t)\big) \tag{6.15}$$

is a rectangular array of six columns, corresponding to the six partial derivatives, and three lines, corresponding to the three components of the three-dimensional vectors, and $\boldsymbol{\alpha}^{[I+1]}(t)$ is a column array, the transpose of which is defined by:

$$\left(\boldsymbol{\alpha}^{[I+1]}\right)^T(t) = \left(\alpha_1^{[I+1]}(t), \alpha_2^{[I+1]}(t), \ldots, \alpha_6^{[I+1]}(t)\right) . \tag{6.16}$$

According to eqn. (5.21) in Chapter 5 the coefficients α_i are expressed as integrals of known functions of time:

$$\boldsymbol{\alpha}^{[I+1]}(t) = \int\limits_{t_0}^{t} \tilde{\mathbf{Z}}^{-1}(t')\, \boldsymbol{F}_p^{[I]}(t')\, dt' , \tag{6.17}$$

where the matrix $\tilde{\mathbf{Z}}$ is the following regular 6×6 matrix

$$\tilde{\mathbf{Z}} = \begin{pmatrix} \boldsymbol{z}_1 & \boldsymbol{z}_2 & \ldots & \boldsymbol{z}_6 \\ \dot{\boldsymbol{z}}_1 & \dot{\boldsymbol{z}}_2 & \ldots & \dot{\boldsymbol{z}}_6 \end{pmatrix} , \tag{6.18}$$

and where the one-dimensional array $\boldsymbol{F}_p^{[I]}$ is given by

$$\left(\boldsymbol{F}_p^{[I]}\right)^T = \left(\mathbf{0}^T, \left(\delta \tilde{\boldsymbol{f}}^{[I]}\right)^T\right) , \tag{6.19}$$

where $\mathbf{0}$ is the column-array of three zero elements.

The solution of the equations of motion in the form (6.12) is in many aspects equivalent to the solution of the differential equations for the orbital elements to be discussed now. The equivalence is in particular given regarding the efficiency and the calculation of perturbations in increasing orders. The solution of the initial value problem (6.12) for $I = 0$ corresponds to the perturbations of first order when integrating the equations of motion for the osculating elements (the concept of first and higher order perturbations will be introduced below).

6.3 Gaussian Perturbation Equations

6.3.1 General Form of the Equations

The concept of osculating elements, as represented by eqn. (4.69), assigns one set of osculating orbital elements to every epoch t via the position and velocity vectors $\boldsymbol{r}(t)$ and $\dot{\boldsymbol{r}}(t)$. There is a one-to-one relationship between the osculating elements of epoch t and the corresponding state vector. The transformation equations between the two sets of functions are those of the two-body problem.

Let

$$I(t) \in \{a(t), e(t), i(t), \Omega(t), \omega(t), T_0(t)\} \tag{6.20}$$

be an arbitrary osculating element. When referring to particular elements the following equivalence will be used:

$$\{I_1(t), I_2(t), I_3(t), I_4(t), I_5(t), I_6(t)\} = \{a(t), e(t), i(t), \Omega(t), \omega(t), T_0(t)\} \ . \tag{6.21}$$

The definition (4.69) implies that each osculating element may be written in the form

$$I(t) \overset{\text{def}}{=} I\left(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)\right) \ . \tag{6.22}$$

The time enters only implicitly in this equation via the time-dependence of the state vector.

The differential equation for the element $I(t)$ is obtained by taking the total derivative of eqn. (6.22) w.r.t. the time t. This is done by applying the well-known chain-rule of elementary calculus:

$$\dot{I} = \sum_{l=1}^{3} \left\{ \frac{\partial I}{\partial r_l}\, \dot{r}_l + \frac{\partial I}{\partial \dot{r}_l}\, \ddot{r}_l \right\} = \nabla_r I \cdot \dot{\boldsymbol{r}} \, + \, \nabla_v I \cdot \ddot{\boldsymbol{r}} \ , \tag{6.23}$$

where $\nabla_r I$ designates the position-, $\nabla_v I$ the velocity-gradient of the orbital element I.

Equation (6.23) may be further modified by replacing the second time derivative of the position vector on the right-hand side by the right-hand side of the original differential equation system (6.1):

$$\dot{I} = \nabla_r I \cdot \dot{\boldsymbol{r}} \, + \, \nabla_v I \cdot \left\{ -\, \mu \, \frac{\boldsymbol{r}}{r^3} + \delta \boldsymbol{f} \right\} \ .$$

Because I is a constant of integration of the two-body problem one may conclude that

$$\dot{I} = \nabla_v I \cdot \delta \boldsymbol{f} \ , \tag{6.24}$$

which is why the complete differential equation system for the entire set of osculating orbital elements assumes the amazingly simple form

$$\dot{I}_k = \nabla_v I_k \cdot \delta \boldsymbol{f} \ , \quad k = 1, 2, \ldots, 6 \ . \tag{6.25}$$

Equations (6.25) are the so-called *Gaussian perturbation equations.* They are not yet in a very useful form – but all that has to be done is to calculate the scalar products on the right-hand sides of eqns. (6.25).

From the mathematical point of view eqns. (6.25) are an explicit, non-linear system of six first-order differential equations. The system is equivalent to the second-order system (6.1). Note, that the scalar products are small quantities of the first order because the perturbation term $\delta \boldsymbol{f}$ is a small quantity of the first order.

If the perturbation term $\delta \boldsymbol{f}$ on the right-hand sides of eqns. (6.25) may be written as gradient of a scalar function R, the so-called *perturbation function*, the perturbation equations may be given the elegant form

$$\dot{I}_k = \nabla_v I_k \cdot \nabla_r R \ , \quad k = 1, 2, \ldots, 6 \ . \tag{6.26}$$

The equations (6.25) and (6.26) are called *perturbation equations in the Gaussian form.* In the form (6.25) they are applicable to a very broad class of perturbations. Gauss derived the perturbation equations by starting from the Lagrangian formulation (which will be given below).

The perturbation equations have a very simple structure. The remaining problem only resides in the fact that the formulas of the two-body problem have to be used explicitly to compute the gradients of the orbital elements w.r.t. the velocity components. The procedure is illustrated for the semi-major axis a in the case of an elliptic motion.

6.3.2 The Equation for the Semi-major Axis *a*

Equation (4.20) defines the energy of the two-body motion. Considering only elliptic motion, we reorder this equation to give the semi-major axis a as a function of the state vectors:

$$\frac{\mu}{a} = \frac{2\,\mu}{r} - \dot{\boldsymbol{r}}^2 \ . \tag{6.27}$$

Taking on both sides the gradient w.r.t. the velocity components we obtain

$$\nabla_v \left(\frac{\mu}{a} \right) = -\,\frac{\mu}{a^2}\,\nabla_v a = -\,2\,\dot{\boldsymbol{r}} \ , \tag{6.28}$$

resulting in

$$\nabla_v a = \frac{2\,a^2}{\mu}\,\dot{\boldsymbol{r}} \ . \tag{6.29}$$

The perturbation equation for the semi-major axis a in the Gaussian form thus reads as

$$\dot{a} = \frac{2\,a^2}{\mu}\,\dot{\boldsymbol{r}} \cdot \delta \boldsymbol{f} \ . \tag{6.30}$$

The same pattern will be used to calculate the gradients of the other orbital elements. Observe that “only” the formulae of the two-body problem as provided in Chapter 4 are required for that purpose.

6.3.3 The Gaussian Equations in Terms of Vectors $\boldsymbol{h}$, $\boldsymbol{q}$

Perturbation Equations for $\boldsymbol{h}$ and $\boldsymbol{q}$. The above derivation of the Gaussian perturbation equation for the semi-major axis a has (hopefully) illustrated that the perturbation equations for the classical orbital elements may be obtained in principle in a straightforward, perhaps though not always in a technically simple way. A relatively simple method to derive all of the six classical elements results, if systematic use is made of the angular momentum vector $\boldsymbol{h}$ (see definition (4.3)) and the Laplacian vector $\boldsymbol{q}$ (see definition (4.27)).

From the developments in Chapter 4 we know that the vectors $\boldsymbol{h}$ and $\boldsymbol{q}$ are first integrals and that they are expressible as simple functions of vectors $\boldsymbol{r}$ and $\dot{\boldsymbol{r}}$. Two vectors with 3 components each, this even promises to take care of the six independent first integrals of the two-body problem. This is not the case, however: The two vectors are not independent, but related by

$$\boldsymbol{h} \cdot \boldsymbol{q} = \mathbf{0} \ , \tag{6.31}$$

because the vector $\boldsymbol{h}$ must always be perpendicular to the instantaneous orbital plane, whereas the vector $\boldsymbol{q}$ always must lie in this plane, by definition.

The five elements p, Ω, i, e, and ω may be easily derived from the two vectors $\boldsymbol{h}$ and $\boldsymbol{q}$ after having solved their perturbation equations (plus the one for the time T_0 of pericenter passage). Therefore, in this straightforward and formally very simple approach, seven instead of six perturbation equations have to be considered. A reduction to six would be rather easily achieved by making use of relation (6.31).

Alternatively, the perturbation equations for the classical elements may be derived using the corresponding equations for $\boldsymbol{h}$ and $\boldsymbol{q}$. This is the more attractive way, because we avoid solving for more than the six independent functions of the problem. This approach is followed subsequently.

Before providing the perturbation equations for $\boldsymbol{h}$ and $\boldsymbol{q}$ we recapitulate, for convenience, the relevant relations (compare eqns. (4.18), (4.5), (4.31), and (4.27)) of the two-body problem:

$$\begin{aligned}
p &= \frac{h^2}{\mu} = \frac{\boldsymbol{h}\cdot\boldsymbol{h}}{\mu} \\
\Omega &= \arctan\left(\frac{h_1}{-h_2}\right) = \arctan\left(\frac{\boldsymbol{h}\cdot\boldsymbol{e}_1}{-\boldsymbol{h}\cdot\boldsymbol{e}_2}\right) \\
i &= \arccos\left(\frac{h_3}{|\boldsymbol{h}|}\right) = \arccos\left(\frac{\boldsymbol{h}\cdot\boldsymbol{e}_3}{|\boldsymbol{h}|}\right) \\
e &= \frac{q}{\mu} = \frac{1}{\mu}\sqrt{\boldsymbol{q}\cdot\boldsymbol{q}} \\
\omega &= \arccos\left\{\frac{\boldsymbol{e}_\Omega\cdot\boldsymbol{q}}{q}\right\} \ ,
\end{aligned} \tag{6.32}$$

where the vectors $\boldsymbol{e}_i$, $i = 1, 2, 3$, are the unit vectors coinciding with the three axes of the (quasi-)inertial Cartesian coordinate system used. The fifth and last equation follows from calculating the scalar product of the Laplacian vector (4.33) and the unit vector $\boldsymbol{e}_\Omega$ (the unit vector of the first axis of the coordinate system Ω of Table 4.3) pointing to the ascending node. Note that the vector $\boldsymbol{e}_\Omega$ was expressed by the vectors $\boldsymbol{h}$ and $\boldsymbol{e}_3$:

$$\boldsymbol{e}_\Omega = \frac{\boldsymbol{e}_3\times\boldsymbol{h}}{|\boldsymbol{e}_3\times\boldsymbol{h}|} = \frac{\boldsymbol{e}_3\times\boldsymbol{h}}{h\,\sin i} \ . \tag{6.33}$$

The perturbation equations for vectors $\boldsymbol{h}$ and $\boldsymbol{q}$ are obtained by formally taking the time derivative of the corresponding defining equations (4.3) and (4.27) and by observing that the two vectors are first integrals in the case of the two-body motion:

$$\dot{\boldsymbol{h}} = \boldsymbol{r}\times\delta\boldsymbol{f} \tag{6.34}$$

and

$$\dot{\boldsymbol{q}} = \delta\boldsymbol{f}\times\boldsymbol{h} + \dot{\boldsymbol{r}}\times\dot{\boldsymbol{h}} \ . \tag{6.35}$$

The six equations (6.34) and (6.35) hold for all orbit types. Together with the equation for T_0 (not yet provided) and the relation (6.31) they fully describe the perturbation problem. The above equations are well known in literature (see, e.g., [94]). Often, they are cited together with the equation for the "energy" (as defined by eqn. (4.9), which is equivalent with the equation for the semi-major axis a for elliptic and hyperbolic orbits). This is not really necessary, because the equation for a may be extracted from the eqns. (6.34) and (6.35).

Equations for $\boldsymbol{p}$, $\boldsymbol{i}$, $\boldsymbol{\Omega}$, $\boldsymbol{e}$, and $\boldsymbol{\omega}$. The perturbation equation for the semi-latus rectum p is obtained by taking the time derivative of the first of eqns. (6.32)

$$\dot{p} = \frac{2}{\mu}\ \boldsymbol{h}\cdot\dot{\boldsymbol{h}} \ . \tag{6.36}$$

The perturbation equation for the eccentricity is obtained by taking the time derivative of the fourth of eqns. (6.32)

$$\dot{e} = \frac{1}{\mu q} \, \boldsymbol{q} \cdot \dot{\boldsymbol{q}} \; . \tag{6.37}$$

As the semi-latus rectum p, the eccentricity e, and the semi-major axis a are related by the equations provided in Table 4.1, we may also give the equation for the semi-major axis a from the above two equations:

$$\dot{a} = \begin{cases} \dfrac{1}{1-e^2} \left(\dot{p} \, + \, 2\, a \, e \, \dot{e}\right) \, , & e < 1 \\ \dfrac{1}{e^2-1} \left(\dot{p} \, - \, 2\, a \, e \, \dot{e}\right) \, , & e > 1 \; . \end{cases} \tag{6.38}$$

From the second and third of eqns. (6.32) we obtain the perturbation equations for the elements i and Ω:

$$\begin{aligned} \frac{di}{dt} &= - \frac{1}{\sqrt{h_1^2 + h_2^2}} \left\{ \dot{h}_3 - \frac{h_3}{h} \, \dot{h} \right\} \\ \dot{\Omega} &= \frac{1}{h_1^2 + h_2^2} \left\{ h_1 \, \dot{h}_2 - h_2 \, \dot{h}_1 \right\} \; . \end{aligned} \tag{6.39}$$

Using the relations

$$\boldsymbol{e}_\Omega = \begin{pmatrix} \cos \Omega \\ \sin \Omega \\ 0 \end{pmatrix} \; , \quad \dot{\boldsymbol{e}}_\Omega = \dot{\Omega} \begin{pmatrix} -\sin \Omega \\ \cos \Omega \\ 0 \end{pmatrix} \; , \quad |\, \boldsymbol{e}_\Omega \cdot \boldsymbol{q} \,| = q \cos \omega \; , \tag{6.40}$$

one obtains the perturbation equation for the argument of pericenter

$$\omega = \arccos \left\{ \frac{\boldsymbol{q} \cdot \boldsymbol{e}_\Omega}{q} \right\} \tag{6.41}$$

by taking the time derivative of the above equation

$$\dot{\omega} = \frac{1}{q \, \sin \omega} \left\{ \cos \omega \, \dot{q} \, - \, \boldsymbol{e}_\Omega \cdot \dot{\boldsymbol{q}} \right\} \, - \, \cos i \, \dot{\Omega} \; . \tag{6.42}$$

Perturbation Equation for $\boldsymbol{T_0}$. The equation for the time T_0 of pericenter passage must be obtained by taking the time derivative of the solution of the equation (4.35) for the argument of latitude u (or for the true anomaly v). The concrete solutions were found to depend on whether the eccentricity e is $e < 1$ (ellipse), $e = 1$ (parabola), or $e > 1$ (hyperbola). The perturbation equation for T_0 may thus be different for different orbit types, as well. We will see below that the differences are of a minor nature – as a matter of fact they might be avoided, if one would formally define the hyperbolic semi-major axis to be negative. The parabolic case usually does not matter when dealing with perturbation equations. This is why only the elliptic and the hyperbolic case are considered, subsequently.

Let us first handle the case of the elliptic orbit, where $e < 1$. For this purpose we start from Kepler's equation

$$E = \sqrt{\frac{\mu}{a^3}}\,(t - T_0) \,+\, e\,\sin E \ , \tag{6.43}$$

solving the equation for the true anomaly, when using the eccentric anomaly E as an auxiliary angle (see Figure 4.4).

Taking the time derivative of Kepler's equation one obtains after a slight reordering of the terms (and by making use of the fact that $n^2\,a^3 = \mu$)

$$\dot E\,(1 - e\,\cos E) = -\frac{3\,n}{2\,a}\,(t - T_0)\,\dot a \,+\, n\left(1 - \dot T_0\right) \,+\, \dot e\,\sin E \ . \tag{6.44}$$

Making furthermore use of the fact that according to eqn. (4.53) $r = a\,(1 - e\,\cos E)$ we obtain the following equation for the time of pericenter passage:

$$\begin{aligned} \dot T_0 &= \frac{\sin E}{n}\,\dot e \,-\, \frac{3}{2\,a}\,(t - T_0)\,\dot a \,+\, 1 \,-\, \frac{r}{a\,n}\,\dot E \\ &= \frac{r\,\sin v}{a\,n\,\sqrt{1-e^2}}\,\dot e \,-\, \frac{3}{2\,a}\,(t - T_0)\,\dot a \,+\, 1 \,-\, \frac{r}{a\,n}\,\dot E \ . \end{aligned} \tag{6.45}$$

The time derivative of the eccentric anomaly is obtained from eqn. (4.63), i.e., from $r = a\,(1 - e\,\cos E)$:

$$\begin{aligned} \dot r \qquad\qquad &= \dot a\,(1 - e\,\cos E) \,-\, a\,\dot e\,\cos E \,+\, a\,e\,\sin E\,\dot E \\ \frac{n\,a}{\sqrt{1-e^2}}\,e\,\sin v &= \frac{r}{a}\,\dot a - (a\,e + r\,\cos v)\,\dot e \,+\, \frac{r\,e\,\sin v}{\sqrt{1-e^2}}\,\dot E \ , \end{aligned} \tag{6.46}$$

where use was made of eqn. (4.64) for the transformation on the left-hand side and of eqns. (4.62) for the transformations on the right-hand side of the above equations. The time derivative of the eccentric anomaly E thus reads as follows:

$$\dot E = \frac{n\,a}{r} + \frac{\sqrt{1-e^2}}{e\,\sin v}\left\{\frac{e + \cos v}{1-e^2}\,\dot e - \frac{\dot a}{a}\right\} \ . \tag{6.47}$$

Introducing this latter result into the equation (6.45) for the time of pericenter passage we obtain eventually

$$\dot T_0 = -\frac{\sqrt{1-e^2}}{e\,n\,\sin v}\left\{\cos v\,\dot e - \frac{r}{a^2}\,\dot a\right\} \,-\, \frac{3}{2\,a}\,(t - T_0)\,\dot a \ , \quad e < 1 \ . \tag{6.48}$$

Note that the perturbation equation for the time of pericenter passage is a linear combination of the perturbation equations in the semi-major axis a and the eccentricity e, where the coefficient of $\dot a$ is explicitly time-dependence.

When dealing with hyperbolic orbits one has to depart from the equivalent of Kepler's equation (see Table 4.2)

$$e\,\sinh F - F = \sqrt{\frac{\mu}{a^3}}\,(t-T_0)\ , \tag{6.49}$$

which solves the equation (4.35) of the true anomaly using the auxiliary angle F. Taking the time derivative of the above equation results in

$$\begin{aligned}(e\,\cosh F - 1)\,\dot{F} &= -\,\frac{3\,n}{2\,a}\,(t-T_0)\,\dot{a}\ +\ n\big(1-\dot{T}_0\big)\ -\ \sinh F\,\dot{e}\\ \frac{r}{a}\,\dot{F} &= -\,\frac{3\,n}{2\,a}\,(t-T_0)\,\dot{a}\ +\ n\big(1-\dot{T}_0\big)\ -\ \sinh F\,\dot{e}\ .\end{aligned} \tag{6.50}$$

Making use of eqns. (4.52, 4.54) we obtain the following equation for the time of pericenter passage:

$$\begin{aligned}\dot{T}_0 &= -\,\frac{\sinh F}{n}\,\dot{e}\ -\ \frac{3}{2a}\,(t-T_0)\,\dot{a}\ +\ 1\ -\ \frac{r}{a\,n}\,\dot{F}\\ &= -\,\frac{r\,\sin v}{a\,n\,\sqrt{e^2-1}}\,\dot{e}\ -\ \frac{3}{2\,a}\,(t-T_0)\,\dot{a}\ +\ 1\ -\ \frac{r}{a\,n}\,\dot{F}\ .\end{aligned} \tag{6.51}$$

The resemblance of this result with the corresponding one (6.45) for the elliptic motion is striking.

The time derivative of the angle F is obtained by taking the time derivative of eqn. (4.54). Observe that the left-hand side (as expressed in terms of the true anomaly v) is identical for both, the elliptic and the hyperbolic orbit – except for replacing $1-e^2$ by e^2-1:

$$\dot{r} = \frac{n\,a}{\sqrt{e^2-1}}\,e\,\sin v = \frac{r}{a}\,\dot{a}\ +\ a\,\cosh F\,\dot{e}\ +\ a\,e\,\sinh F\,\dot{F}\ . \tag{6.52}$$

Making use of eqns. (4.52) it is a straightforward procedure to derive the expression for the time derivative of F:

$$\dot{F} = \frac{n\,a}{r}\ -\ \frac{\sqrt{e^2-1}}{e\,\sin v}\left\{\frac{e+\cos v}{e^2-1}\,\dot{e}+\frac{\dot{a}}{a}\right\}\ . \tag{6.53}$$

Observe the similarity of the above expression with the corresponding derivative (6.47) for the elliptic motion.

Introducing the result (6.53) into the equation (6.51) leads to the final result for hyperbolic orbits

$$\dot{T}_0 = \frac{\sqrt{e^2-1}}{e\,n\,\sin v}\left\{\cos v\,\dot{e}+\frac{r}{a^2}\,\dot{a}\right\}\ -\ \frac{3}{2\,a}\,(t-T_0)\,\dot{a}\ ,\quad e>1\ , \tag{6.54}$$

which is, as expected, rather similar to the corresponding result (6.48) of the elliptic motion.

Summary. What was achieved in this section? Starting from the perturbation equations (6.34, 6.35) for the angular momentum vector $\boldsymbol{h}$ and the Laplacian vector $\boldsymbol{q}$ the perturbation equations for the classical osculating orbital elements were derived. Five equations, namely those for p, e, i, Ω, and ω, hold for all possible orbit types. The equation for the time T_0 of pericenter passage is case sensitive. Below, we use the vector notation, as opposed to the coordinate notation used so far by, e.g., using the identity $h_3 = \boldsymbol{h} \cdot \boldsymbol{e}_3$, the latter vector being the unit vector defining the third axis of the (quasi-) inertial Cartesian coordinate system. The six perturbation equations for the classical osculating elements, expressed in terms of the vectors $\boldsymbol{h}$ and $\boldsymbol{q}$, are:

$$\begin{aligned}
\dot{p} &= \frac{2}{\mu}\, \boldsymbol{h} \cdot \dot{\boldsymbol{h}} \\
\frac{di}{dt} &= -\frac{1}{\sqrt{h_1^2 + h_2^2}} \left\{ \dot{\boldsymbol{h}} \cdot \boldsymbol{e}_3 - \frac{h_3}{h}\, \dot{h} \right\} \\
\dot{\Omega} &= \frac{1}{h_1^2 + h_2^2} \left\{ (\boldsymbol{h} \times \dot{\boldsymbol{h}}) \cdot \boldsymbol{e}_3 \right\} \\
\dot{e} &= \frac{1}{\mu\, q}\, \boldsymbol{q} \cdot \dot{\boldsymbol{q}} \\
\dot{\omega} &= \frac{1}{q\, \sin\omega} \{ \cos\omega\, \dot{q} - \boldsymbol{e}_\Omega \cdot \dot{\boldsymbol{q}} \} - \cos i\, \dot{\Omega} \\
\dot{T}_0 &= -\frac{3}{2\,a}\,(t - T_0)\,\dot{a} + \frac{\sqrt{|1-e^2|}}{e\,n\,\sin v} \cdot \begin{cases} \left[-\cos v\, \dot{e} + \dfrac{r}{a^2}\, \dot{a} \right], & e < 1 \\ \left[+\cos v\, \dot{e} + \dfrac{r}{a^2}\, \dot{a} \right], & e > 1 \end{cases} .
\end{aligned} \tag{6.55}$$

The above equations actually form a self-contained differential equation system for the orbital elements p, i, Ω, e, ω, and T_0. They may be used to develop a computer program, where the vectors of the scalar products on the right-hand sides, e.g., in $\boldsymbol{h} \cdot \dot{\boldsymbol{h}}$, have to be replaced by the corresponding right-hand sides of the defining equations (4.3), (4.27) and their time derivatives (6.34), (6.35).

6.3.4 Gaussian Perturbation Equations in Standard Form

Equations (6.55) may be brought into the standard form (6.26) in essence by repeatedly using the theorem

$$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{c} \cdot (\boldsymbol{a} \times \boldsymbol{b}) \tag{6.56}$$

of vector algebra.

The results may be written in different levels of explicitness. On one hand, one would like to retain a general vector notation (the choice of the mathematician), on the other hand one would like to use, to the extent possible, the well-established formulae of the two-body problem to guarantee that the equations may be used easily by the practitioner. In our sketch of the transformation of equations (6.55) we will first derive the general vector notation, then provide the (hopefully) optimal version for application.

Equation for the Semi-latus Rectum *p*. On the right-hand side of the first of eqns. (6.55) the scalar product may be transformed as follows:

$$\boldsymbol{h} \cdot \dot{\boldsymbol{h}} = \boldsymbol{h} \cdot (\boldsymbol{r} \times \delta\boldsymbol{f}) = (\boldsymbol{h} \times \boldsymbol{r}) \cdot \delta\boldsymbol{f} \ . \tag{6.57}$$

The equation for p may thus be written as

$$\dot{p} = \frac{2}{\mu}\ (\boldsymbol{h} \times \boldsymbol{r}) \cdot \delta\boldsymbol{f} = 2\sqrt{\frac{p}{\mu}}\ r\left(\boldsymbol{e}_{\mathcal{R}_2} \cdot \delta\boldsymbol{f}\right)\ , \tag{6.58}$$

where $\boldsymbol{e}_{\mathcal{R}_2} \overset{\text{def}}{=} \frac{\boldsymbol{h}\times\boldsymbol{r}}{h\,r}$ is the unit vector lying in the second axis of the orbital system (see Table 4.3). The vector is perpendicular to the position vector $\boldsymbol{r}$ and points (more or less) into the direction of motion.

The above equation says that p is only affected, if the perturbing acceleration $\delta\boldsymbol{f}$ has a component collinear with vector $\boldsymbol{e}_{\mathcal{R}_2}$. One might wish to use another form for the above equation. As the vector $\boldsymbol{e}_{\mathcal{R}_2}$ lies in the orbital plane, it can be represented by the first two unit vectors $\boldsymbol{e}_{\Pi_1}$ (pointing to the pericenter) and $\boldsymbol{e}_{\Pi_2}$ of the orbital coordinate system. The transformation is given by (see Table 4.3)

$$\boldsymbol{e}_{\mathcal{R}_2} = \cos\left(\frac{\pi}{2} + v\right)\ \boldsymbol{e}_{\Pi_1}\ +\ \sin\left(\frac{\pi}{2} + v\right)\ \boldsymbol{e}_{\Pi_2} = -\ \sin v\ \boldsymbol{e}_{\Pi_1}\ +\ \cos v\ \boldsymbol{e}_{\Pi_2}\ . \tag{6.59}$$

The perturbation equation for the semi-latus rectum p therefore also may be given the form

$$\dot{p} = 2\sqrt{\frac{p}{\mu}}\ r\left\{-\sin v\ \boldsymbol{e}_{\Pi_1}\ +\ \cos v\ \boldsymbol{e}_{\Pi_2}\right\} \cdot \delta\boldsymbol{f}\ . \tag{6.60}$$

Equations (6.58) and (6.60) are mathematically equivalent. From the physical point of view one would probably prefer the first representation, because it shows immediately that only the component of $\delta\boldsymbol{f}$ in $\boldsymbol{e}_{\mathcal{R}_2}$-direction matters. In practice it may, however, be better to use the second representation, because the unit vector of the orbital system usually have to be calculated anyway.

Equation for Inclination *i*. In the second of eqns. (6.55) we may transform the bracket $\{\ldots\}$ (using eqn. (6.56)) on the right-hand side as follows:

$$\dot{\boldsymbol{h}} \cdot \boldsymbol{e}_3\ -\ \frac{h_3}{h}\,\dot{h} = (\boldsymbol{r} \times \delta\boldsymbol{f}) \cdot \boldsymbol{e}_3\ -\ \frac{h_3}{h^2}\ (\boldsymbol{h} \times \boldsymbol{r}) \cdot \delta\boldsymbol{f}\ . \tag{6.61}$$

The second term could be taken over from eqn. (6.57), the first one may now be developed using eqn. (6.56). The result reads as

$$\dot{\boldsymbol{h}} \cdot \boldsymbol{e}_3 - \frac{h_3}{h} \dot{h} = \left\{ (\boldsymbol{e}_3 \times \boldsymbol{r}) - \frac{h_3}{h^2} (\boldsymbol{h} \times \boldsymbol{r}) \right\} \cdot \delta\boldsymbol{f} \ . \tag{6.62}$$

Making use of the relations

$$h_3 = h \cos i \ , \quad \sqrt{h_1^2 + h_2^2} = h \sin i \ , \tag{6.63}$$

the perturbation equation for the inclination i may be written in the form

$$\frac{di}{dt} = -\frac{1}{h \sin i} \left\{ (\boldsymbol{e}_3 - \cos i \ \boldsymbol{e}_{\Pi_3}) \times \boldsymbol{r} \right\} \cdot \delta\boldsymbol{f} \overset{\text{def}}{=} + \frac{1}{h \sin i} (\boldsymbol{r} \times \boldsymbol{x}) \cdot \delta\boldsymbol{f} \ , \tag{6.64}$$

where $\boldsymbol{e}_{\Pi_3}$, the third unit vector of the orbital coordinate system, is normal to the orbital plane. The vector $\boldsymbol{x}$, defined above, lies in the orbital plane, points to the point of maximum elevation above the reference plane (at argument of latitude $u = 90°$) and has the length $|\boldsymbol{x}| = \sin i$. The vector product $\boldsymbol{x} \times \boldsymbol{r}$ therefore is collinear with vector $\boldsymbol{h}$. The non-zero (third) coordinate in the orbital system is given by

$$|\boldsymbol{x}| \, |\boldsymbol{h}| \sin\left(\frac{\pi}{2} - u\right) = \sin i \sqrt{\mu p} \ \cos u \ , \tag{6.65}$$

which is why the perturbation equation for the inclination i also may be brought into the form

$$\frac{di}{dt} = \frac{r \cos u}{n a^2 \sqrt{|1 - e^2|}} \ \boldsymbol{e}_{\Pi_3} \cdot \delta\boldsymbol{f} \ . \tag{6.66}$$

Equations (6.64) and (6.66) are equivalent, where the second version undoubtedly is preferable from the practical point of view.

Equation for Node *Ω*. In order to obtain the perturbation equation for the element Ω (third of eqns. (6.55)) we have to transform the bracket on the right-hand side of this equation as follows:

$$(\boldsymbol{h} \times \dot{\boldsymbol{h}}) \cdot \boldsymbol{e}_3 = (\boldsymbol{e}_3 \times \boldsymbol{h}) \cdot \dot{\boldsymbol{h}} = (\boldsymbol{e}_3 \times \boldsymbol{h}) \cdot (\boldsymbol{r} \times \delta\boldsymbol{f}) = \big((\boldsymbol{e}_3 \times \boldsymbol{h}) \times \boldsymbol{r} \big) \cdot \delta\boldsymbol{f} \ . \tag{6.67}$$

Note that $\boldsymbol{e}_3 \times \boldsymbol{h} = h \sin i \ \boldsymbol{e}_\Omega$, which is why

$$\big((\boldsymbol{e}_3 \times \boldsymbol{h}) \times \boldsymbol{r} \big) = h \sin i \ r \sin u \ \frac{\boldsymbol{h}}{h} \ . \tag{6.68}$$

As

$$h_1^2 + h_2^2 = h^2 \sin^2 i \quad \text{and} \quad \boldsymbol{e}_{\Pi_3} = \frac{\boldsymbol{h}}{h} \ , \tag{6.69}$$

the perturbation equation for Ω may be written as

$$\dot{\Omega} = \frac{1}{h^2 \sin^2 i} \big((\boldsymbol{e}_3 \times \boldsymbol{h}) \times \boldsymbol{r} \big) \cdot \delta\boldsymbol{f} = \frac{r \sin u}{n a^2 \sin i \sqrt{|1 - e^2|}} \ (\boldsymbol{e}_{\Pi_3} \cdot \delta\boldsymbol{f}) \ . \tag{6.70}$$

As in the case of the other elements, two versions were provided to express the perturbation equation for the node Ω. Both are in the standard form (6.26). The second version is more useful from the practical point of view.

Equation for Eccentricity *e*. In order to obtain the perturbation equation for the eccentricity e we have to calculate the scalar product

$$\boldsymbol{q}\cdot\dot{\boldsymbol{q}}=\boldsymbol{q}\cdot\big\{\delta\boldsymbol{f}\times\boldsymbol{h}+\dot{\boldsymbol{r}}\times(\boldsymbol{r}\times\delta\boldsymbol{f})\big\}=\big\{\boldsymbol{h}\times\boldsymbol{q}-\boldsymbol{r}\times(\boldsymbol{q}\times\dot{\boldsymbol{r}})\big\}\cdot\delta\boldsymbol{f}\ . \qquad (6.71)$$

The vector $\boldsymbol{h}\times\boldsymbol{q}$ is of length hq, lies in the orbital plane and is perpendicular to $\boldsymbol{q}$, thus collinear with the second axis $\boldsymbol{e}_{\Pi_2}$ of the orbital system Π (see Table 4.3). Making use of the theorem (4.24) of vector analysis the product $\boldsymbol{r}\times(\boldsymbol{q}\times\dot{\boldsymbol{r}})$ may be written as a linear combination of the vectors $\boldsymbol{e}_{Pi_1}$ and $\dot{\boldsymbol{r}}$. A first version for the eccentricity thus may be written as

$$\dot{e}=\sqrt{\frac{p}{\mu}}\left\{\boldsymbol{e}_{\Pi_2}-\frac{\boldsymbol{r}\cdot\dot{\boldsymbol{r}}}{h}\,\boldsymbol{e}_{\Pi_1}+\frac{\boldsymbol{r}\cdot\boldsymbol{e}_{\Pi_1}}{h}\,\dot{\boldsymbol{r}}\right\}\cdot\delta\boldsymbol{f}\ . \qquad (6.72)$$

The three vectors in the parentheses $\{\ldots\}$ of the above expression are lying in the instantaneous orbital plane. It is therefore possible to write one of the two vectors as a linear combination of the two others. It seems reasonable to represent the third vector $\dot{\boldsymbol{r}}$ by $\boldsymbol{e}_{\Pi_1}$ and $\boldsymbol{e}_{\Pi_2}$. The transformation equations (4.62) and (4.64) are used for this purpose. In vectorial notation the transformation reads as

$$\frac{1}{h}\,(\boldsymbol{r}\cdot\boldsymbol{e}_{\Pi_1})\,\dot{\boldsymbol{r}}=\frac{r}{p}\,\cos v\,\big\{-\sin v\ \boldsymbol{e}_{\Pi_1}+(e+\cos v)\ \boldsymbol{e}_{\Pi_2}\big\}\ . \qquad (6.73)$$

Using the same transformation equations we may also transform the scalar product

$$\frac{1}{h}\,(\boldsymbol{r}\cdot\dot{\boldsymbol{r}})=\frac{r}{p}\,e\,\sin v\ . \qquad (6.74)$$

The perturbation equation for the eccentricity e thus may be written as

$$\dot{e}=\sqrt{\frac{p}{\mu}}\left\{\boldsymbol{e}_{\Pi_2}+\frac{r}{p}\,(e+\cos v)\,(-\sin v\ \boldsymbol{e}_{\Pi_1}+\cos v\ \boldsymbol{e}_{\Pi_2})\right\}\cdot\delta\boldsymbol{f}\ . \qquad (6.75)$$

Equation for Argument of the Pericenter *ω*. As the perturbation equation for Ω and $q=\mu e$ were already given above, we only have to calculate the term

$$\boldsymbol{e}_{\Omega}\cdot\dot{\boldsymbol{q}}=\boldsymbol{e}_{\Omega}\,\big\{\delta\boldsymbol{f}\times\boldsymbol{h}+\dot{\boldsymbol{r}}\times(\boldsymbol{r}\times\delta\boldsymbol{f})\big\}\ . \qquad (6.76)$$

Using the theorem (6.56) it is a straightforward matter to show that

$$\boldsymbol{e}_{\Omega}\cdot\dot{\boldsymbol{q}}=\big\{\,(\boldsymbol{h}\times\boldsymbol{e}_{\Omega})+(\boldsymbol{e}_{\Omega}\times\dot{\boldsymbol{r}})\times\boldsymbol{r}\big\}\cdot\delta\boldsymbol{f}\ . \qquad (6.77)$$

The first vector on the right-hand side is of length $|\,\boldsymbol{h}\times\boldsymbol{e}_{\Omega}\,|=h$, lies in the orbital plane and points to the argument of latitude $u=\pi/2$. The second vector lies in the orbital plane, as well, has the "length" of $\sqrt{\mu/p}\,r\,(\cos u+e\cos\omega)$, and is collinear with $\boldsymbol{e}_S$. (Use the equations (4.62) and (4.64) to calculate the double vector product).

It is a straightforward procedure to show that the three terms on the right-hand side of the fifth of eqns. (6.55) may be written as

$$\dot{\omega} = - \sqrt{\frac{p}{\mu}} \, \frac{1}{e} \left\{ \boldsymbol{e}_{\Pi_1} + \frac{r}{p} \, \sin v \, (- \sin v \, \boldsymbol{e}_{\Pi_1} + \cos v \, \boldsymbol{e}_{\Pi_2}) \right\} \cdot \delta \boldsymbol{f} \; - \; \cos i \; \dot{\Omega} \; . \tag{6.78}$$

Time of Pericenter Passage. The time of pericenter passage may be expressed as a linear combination of the equations for the semi-major axis a and the eccentricity e. The differential equation for the eccentricity e is already available in the standard form. In order to express the equation for the time T_0 of pericenter passage in the standard form, we need the equation for the semi-major axis in the same form, as well. This equation may easily be derived using the relation (6.38). On the other hand, we already gave the result for an elliptic orbit in eqn. (6.30). The result is

$$\dot{a} = \frac{2}{n^2 \, a} \begin{cases} + \dot{\boldsymbol{r}} \cdot \delta \boldsymbol{f} \; ; & e < 1 \\ - \dot{\boldsymbol{r}} \cdot \delta \boldsymbol{f} \; ; & e > 1 \end{cases} . \tag{6.79}$$

Observe that we avoided the use of the semi-latus rectum p in the above equations because of the different meanings in the elliptic and hyperbolic case.

As we gave the equation for e in terms of the unit vectors $\boldsymbol{e}_{\Pi_i}$, $i = 1, 2$, we do the same for the semi-major axis a:

$$\dot{a} = \frac{2}{n \sqrt{|1 - e^2|}} \begin{cases} + \, \big(- \sin v \; \boldsymbol{e}_{\Pi_1} + (e + \cos v) \; \boldsymbol{e}_{\Pi_2} \big) \cdot \delta \boldsymbol{f} \; ; & e < 1 \\ - \, \big(- \sin v \; \boldsymbol{e}_{\Pi_1} + (e + \cos v) \; \boldsymbol{e}_{\Pi_2} \big) \cdot \delta \boldsymbol{f} \; ; & e > 1 \end{cases} . \tag{6.80}$$

Introducing eqns. (6.80) and (6.75) into the perturbation equations for the element T_0 (6.48) (for elliptic orbits) and (6.54) (for hyperbolic orbits) we obtain the result

$$\begin{aligned} \dot{T}_0 = - \, \frac{1 - e^2}{e \, a \, n^2 \, \sin v} & \left\{ \left(\cos v - 2 \, \frac{r}{p} \, e \right) \boldsymbol{e}_{\Pi_2} \right. \\ & \left. + \, \frac{r}{p} \, \big[\, (e + \cos v) \cos v - 2 \, \big] \, \boldsymbol{e}_{\mathcal{R}_2} \right\} \cdot \delta \boldsymbol{f} \; - \; \frac{3}{2 \, a} \, (t - T_0) \, \dot{a} \end{aligned} \tag{6.81}$$

for elliptic orbits ($e < 1$) and

$$\begin{aligned} \dot{T}_0 = + \, \frac{e^2 - 1}{e \, a \, n^2 \, \sin v} & \left\{ \left(\cos v - 2 \, \frac{r}{p} \, e \right) \boldsymbol{e}_{\Pi_2} \right. \\ & \left. + \, \frac{r}{p} \, \big[\, (e + \cos v) \cos v - 2 \, \big] \, \boldsymbol{e}_{\mathcal{R}_2} \right\} \cdot \delta \boldsymbol{f} \; - \; \frac{3}{2 \, a} \, (t - T_0) \, \dot{a} \end{aligned} \tag{6.82}$$

for hyperbolic orbits ($e > 1$).

Summary. The perturbation equations for the six classical orbital elements were brought into the standard form (see (6.26)) in this section. From our results the velocity-gradients associated with the individual elements may be easily extracted (where we include for convenience, both, the gradients for the semi-latus rectum p and the semi-major axis a):

$$\begin{aligned}
\nabla_v\, p &= 2\sqrt{\frac{p}{\mu}}\; r\; \boldsymbol{e}_S \\
\nabla_v\, i &= \frac{r\,\cos u}{n\,a^2\,\sqrt{|1-e^2|}}\; \boldsymbol{e}_{\Pi_3} \\
\nabla_v\, \Omega &= \frac{r\,\sin u}{n\,a^2\,\sin i\,\sqrt{|1-e^2|}}\; \boldsymbol{e}_{\Pi_3} \\
\nabla_v\, e &= \sqrt{\frac{p}{\mu}}\left\{ \boldsymbol{e}_{\Pi_2} + \frac{r}{p}\,(e+\cos v)\,\boldsymbol{e}_S\right\} \\
\nabla_v\, \omega &= \sqrt{\frac{p}{\mu}}\left\{ \boldsymbol{e}_{\Pi_2} + \frac{r}{p}\,(e+\cos v)\,\boldsymbol{e}_S\right\} \\
\nabla_v\, T_0 + \frac{3}{2\,a}\,(t-T_0)\;\nabla_v\, a &= -\,\frac{1-e^2}{e\,a\,n^2\,\sin v}\left\{\left[\cos v - 2\,\frac{r}{p}\,e\right]\boldsymbol{e}_{\Pi_2}\right. \\
&\qquad \left. + \frac{r}{p}\,\big[\,(e+\cos v)\cos v - 2\big]\,\boldsymbol{e}_{\mathcal{R}_2}\right\} \\
\nabla_v\, a &= \pm\,\frac{2}{n\,\sqrt{|1-e^2|}}\,\big\{-\sin v\;\boldsymbol{e}_{\Pi_1} + (e+\cos v)\;\boldsymbol{e}_{\Pi_2}\big\}\;.
\end{aligned}\tag{6.83}$$

Note that the positive sign in the gradient for the semi-major axis holds for elliptic, the negative for hyperbolic orbits. The equation for T_0 holds for both, elliptic and parabolic orbits!

6.3.5 Decompositions of the Perturbation Term

The scalar products on the right-hand sides of the Gaussian perturbation equations (6.83) may be calculated in any coordinate system that might seem convenient. Two particular Cartesian systems, both rotating w.r.t. inertial space in a rather complicated way and both already defined in Table 4.3, prove to be very useful:

- The $\mathcal{R}$-system system, decomposing the perturbing acceleration into a radial component R', a component S' normal to R' in the orbital plane (pointing approximately into the direction of motion), and the out-of-plane component W' normal to the orbital plane. The components refer to the unit vectors $\boldsymbol{e}_{\mathcal{R}_1}$, $\boldsymbol{e}_{\mathcal{R}_2}$, and $\boldsymbol{e}_{\Pi_3}$.
- The $\mathcal{T}$-system, decomposing the perturbing acceleration into a tangential component T' (parallel to the velocity vector $\dot{\boldsymbol{r}}$), a component N' normal

to T' in the orbital plane (and pointing into the ellipse), and the out-of-plane component W' normal to the orbital plane. The unit vectors of this Cartesian coordinate system shall be denoted by $\boldsymbol{e}_T$, $\boldsymbol{e}_N$, and $\boldsymbol{e}_{\Pi_3}$.

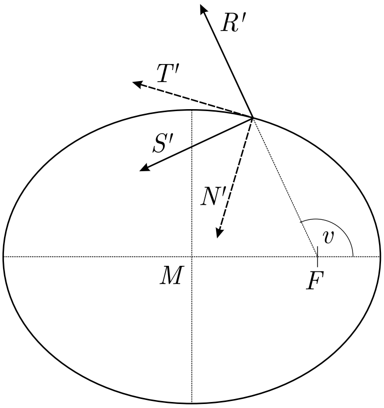


Fig. 6.1. Decomposition of the perturbing acceleration into the radial R', normal to radial in the orbital plane S', and the out-of-plane direction W'; or into the tangential T', normal to tangential in the orbital plane N' and the out-of-plane direction W'

Figure 6.1 illustrates the (R', S', W')- and the (T', N', W')-systems. In the (R', S', W')-system the components of the velocity vector are computed as $\boldsymbol{R}_3(v)\,\dot{\boldsymbol{x}}_\Pi$ using eqn. (4.64). The result is

$$\dot{\boldsymbol{r}}_{\mathcal{R}} = \begin{pmatrix} \dot{r} \\ r\,\dot{u} \\ 0 \end{pmatrix} = \sqrt{\frac{\mu}{p}} \begin{pmatrix} e\,\sin v \\ \frac{p}{r} \\ 0 \end{pmatrix} .$$

In vector notation the same transformation may be written as:

$$\dot{\boldsymbol{r}} = \sqrt{\frac{\mu}{p}} \left\{ e\,\sin v\; \boldsymbol{e}_{\mathcal{R}_1} + \frac{p}{r}\,\boldsymbol{e}_{\mathcal{R}_2} \right\} . \tag{6.84}$$

With this representation of the velocity vector, the perturbation equation for the semi-major axis a in eqns. (6.83) assumes the form

$$\dot{a} = \frac{1}{n\sqrt{1-e^2}} \left(e\,\sin v\, R' + \frac{p}{r}\, S' \right) , \tag{6.85}$$

where an elliptic osculating orbit was assumed.

In the case of the semi-major axis a the decomposition according to the (T', N', W')-system leads to an even simpler result, because in this particular coordinate system the scalar product associated with the element a simply is the product of the absolute value of the velocity with the tangential component T':

$$\dot{\boldsymbol{r}} \cdot \delta \boldsymbol{f} = |\dot{\boldsymbol{r}}| \, T' \ , \tag{6.86}$$

The last of the perturbation equation (6.83) then simply reads as

$$\dot{a} = \frac{2}{n^2 \, a} \, |\dot{\boldsymbol{r}}| \, T' \ , \tag{6.87}$$

where the relation $\mu = n^2 \, a^3$ was used. An elliptic orbit was assumed.

Equation (6.87) illustrates how insight is gained into the structure of a particular perturbation by (a) using the perturbation equations for the orbital elements instead of the original equations in rectangular coordinates and (b) by an appropriate decomposition of the perturbing acceleration. The equation tells that only with a tangential component it is possible to change the semi-major axis; a positive tangential component increases, a negative decreases the semi-major axis. The equation also tells that a tangential acceleration T' of short duration (and of the same size) has maximum effect on the semi-major axis a when applied in the pericenter and minimum effect when applied in the apocenter. Space agencies operating artificial Earth satellites are of course utilizing such basic facts. Manoeuvres intended to change the semi-major axis of a space-craft have to be realized by thrusts in the tangential (along-track) direction. The atmospheric drag acting on an artificial Earth satellite in a circular orbit is an example for a (more or less) constant perturbating force in the along-track direction. It is opposed to the satellite motion, therefore decreases the semi-major axis, and eventually leads to the decay of the satellite.

The complete set of the Gaussian perturbation equations for the (R', S', W')-decomposition is now easily derived from the general representation (6.83):

$$\begin{aligned}
\dot{a} &= \sqrt{\frac{p}{\mu}} \, \frac{2\,a}{1-e^2} \left\{ e \, \sin v \, R' + \frac{p}{r} \, S' \right\} \\
\dot{e} &= \sqrt{\frac{p}{\mu}} \left\{ \sin v \, R' + (\cos v + \cos E) \, S' \right\} \\
\dot{T}_0 &= -\frac{1-e^2}{n^2 \, a \, e} \left\{ \left(\cos v - 2\,e\,\frac{r}{p} \right) R' - \left(1 + \frac{r}{p} \right) \sin v \, S' \right\} - \frac{3}{2\,a} \, (t - T_0) \, \dot{a} \\
\frac{di}{dt} &= \frac{r \, \cos u}{n \, a^2 \, \sqrt{1-e^2}} \, W' \\
\dot{\Omega} &= \frac{r \, \sin u}{n \, a^2 \, \sqrt{1-e^2} \, \sin i} \, W' \\
\dot{\omega} &= \frac{1}{e} \sqrt{\frac{p}{\mu}} \left\{ -\cos v \, R' + \left(1 + \frac{r}{p} \right) \sin v \, S' \right\} - \cos i \, \dot{\Omega} \ ,
\end{aligned} \tag{6.88}$$

where v is the true, E the eccentric anomaly, and $u = \omega + v$ the argument of latitude of the celestial body considered (see Chapter 4). Only the formulae for the case $e < 1$ are reproduced above. The perturbation equations (6.88) are arranged in two groups, the first consisting of the equations for the semi-major axis a (defining the size), the eccentricity e (defining the shape), and the time T_0 of pericenter passage (defining the dynamics) of the orbital motion, the second consisting of the three Eulerian angles i, Ω, and ω defining the orbital plane and the orientation of the conic section within the orbital plane. Observe that the orbital plane can be only changed by an out-of-plane component W'.

The choice of the set a, e, i, Ω, ω and T_0 of osculating elements is by no means unique. Alternatives make in particular sense if the inclination i and/or if the eccentricity e are small. The perturbation equations for such elements are easily obtained by elementary combinations of the set of equations (6.88). We include one alternative for the time of pericenter, by replacing T_0 by the mean anomaly σ_0 referring to the initial epoch t_0. The transformation equation relating σ_0 and T_0 is:

$$\sigma_0 \overset{\text{def}}{=} n\,(t_0 - T_0) \ , \tag{6.89}$$

where according to eqn. (4.41) $n = \sqrt{\frac{\mu}{a^3}}$ is the osculating mean motion of the celestial body. Taking the time derivative of eqn. (6.89) results in the following perturbation equation for the element σ_0:

$$\dot{\sigma}_0 = \frac{1-e^2}{n\,a\,e} \left\{ \left(\cos v - 2\,e\,\frac{r}{p} \right) R' - \left(1 + \frac{r}{p} \right) \sin v\, S' \right\} + \frac{3\,n}{2\,a}\,(t - t_0)\,\dot{a} \ . \tag{6.90}$$

For further reference we conclude this paragraph by the Gaussian perturbation equations for the decomposition T', N', and W'. The equations for i and Ω may be skipped because they contain the W'-component which is common to both decompositions. The result is easily obtained from the general representation (6.83) of the Gaussian equations.

$$\begin{aligned}
\dot{a} &= \frac{2\,a^2}{\mu}\,|\dot{\boldsymbol{r}}|\;T' \\
\dot{e} &= \frac{1}{|\dot{\boldsymbol{r}}|} \left\{ -\frac{r}{a}\,\sin v\,N' + 2\,(\cos v + e)\,T' \right\} \\
\dot{T}_0 &= \frac{\sqrt{1-e^2}}{n\,e\,|\dot{\boldsymbol{r}}|} \left\{ \frac{r}{a}\,\cos v\,N' + 2\left(1 + e^2\,\frac{r}{p} \right) \sin v\,T' \right\} - \frac{3}{2\,a}\,(t - T_0)\,\dot{a} \\
\dot{\omega} &= \frac{1}{e\,|\dot{\boldsymbol{r}}|} \left\{ \left(\frac{r}{p}\,\cos v + e \left(1 + \frac{r}{p} \right) \right) N' + 2\,\sin v\,T' \right\} - \cos i\,\dot{\Omega} \ .
\end{aligned} \tag{6.91}$$

6.4 Lagrange's Planetary Equations

In this section it is assumed that the perturbing acceleration $\delta\boldsymbol{f}$ may be represented as the gradient of the scalar perturbation function R:

$$\delta\boldsymbol{f} \overset{\text{def}}{=} \nabla_r R \ . \tag{6.92}$$

6.4.1 General Form of the Equations

The Gaussian perturbation equations assume the form (6.26):

$$\dot{I}_k = \nabla_v I_k \cdot \nabla_r R \ , \quad k = 1, 2, \ldots, 6 \ . \tag{6.93}$$

Due to the fact that the osculating elements at epoch t may be derived from the state vector referring to the same epoch (and vice versa), the gradient in eqn. (6.26) may be expressed by a linear combination of the gradients (w.r.t. the position vector) of the six osculating elements

$$\nabla_r R = \sum_{j=1}^{6} \frac{\partial R}{\partial I_j} \nabla_r I_j \ , \tag{6.94}$$

where coefficients of the linear combination are the partial derivatives of the perturbation function w.r.t. the corresponding osculating element. Replacing the gradient of the perturbation function on the right-hand side of the Gaussian perturbation equations (6.93) by the right-hand sides of eqn. (6.94) leads to the formally very simple result

$$\dot{I}_k = \sum_{j=1}^{6} \big(\nabla_v I_k \cdot \nabla_r I_j\big) \frac{\partial R}{\partial I_j} \ , \quad k = 1, 2, \ldots, 6 \ . \tag{6.95}$$

The perturbation equations (6.95) represent the time derivative of each orbital element as a linear combination of the perturbation function's partial derivatives w.r.t. all six orbital elements. It may be viewed as a disadvantage of eqns. (6.95) that the sum has to be extended over all six orbital elements (constants of integration).

This situation may be improved by making explicit use of the fact that the perturbation function R does not depend on the velocity components, which is why the velocity-gradient of the perturbation function R is a zero-vector. The analogue of eqn. (6.94) in velocity space therefore reads as

$$\nabla_v R = \sum_{j=1}^{6} \frac{\partial R}{\partial I_j} \nabla_v I_j = \mathbf{0} \ . \tag{6.96}$$

This relationship may be used to eliminate the term with summation index $j = k$ on the right-hand side of the perturbation equations (6.95): We simply multiply the above equation (in the sense of a scalar product) with the vector $\nabla_r I_k$ and subtract the resulting scalar product from eqn. (6.95). The result are the *Lagrange's planetary equations*:

$$\dot{I}_k = \sum_{j=1}^{6} \left(\nabla_v I_k \cdot \nabla_r I_j \; - \; \nabla_v I_j \cdot \nabla_r I_k \right) \frac{\partial R}{\partial I_j} \, , \quad k = 1, 2, \ldots, 6 \; . \tag{6.97}$$

Observe that the term $j = k$ on the right-hand side is zero due to symmetry reasons. It is therefore not necessary to exclude the term number k explicitly in the above sum. The terms $(\ldots)$ are the well known *Poisson brackets* of analytical mechanics, named after Siméon-Denis Poisson (1781–1840). It is remarkable that we arrived at the above representation of the perturbation equations merely by making explicit use of the fact that the function R is the gradient of a scalar function of the position vector (and does not depend on $\dot{\boldsymbol{r}}$).

The following notation for the Poisson bracket referring to the orbital elements k and j will be used subsequently:

$$[I_k, I_j] \overset{\text{def}}{=} \nabla_v I_k \cdot \nabla_r I_j \; - \; \nabla_v I_j \cdot \nabla_r I_k \; . \tag{6.98}$$

The *Lagrange's planetary equations* may thus be given the following elegant form:

$$\dot{I}_k = \sum_{j=1}^{6} [I_k, I_j] \, \frac{\partial R}{\partial I_j} \, , \quad k = 1, 2, \ldots, 6 \; . \tag{6.99}$$

The definition (6.98) of the Poisson bracket implies anti-symmetry

$$[I_k, I_j] = - \, [I_j, I_k] \quad \text{and therefore} \quad [I_k, I_k] = 0 \quad j, k = 1, 2, \ldots, 6 \; , \tag{6.100}$$

which is why only 15 out of the 36 Poisson brackets are independent and have to be computed.

Lagrange's planetary equations (6.99) are not yet very useful in this form, but all that remains to be done is the explicit computation of the Poisson brackets. This will be done in section 6.4.3, where we will see that only five out of the 15 independent Poisson brackets are different from zero.

The general form (6.99) of the equations of motion is useable for a broad class of problems: If a dynamical system may be written as a perturbation problem and if the perturbation term is a position-gradient and does not depend on the velocity, then the perturbation equations may be written in the form (6.99). The problem-dependent part only resides in the explicit computation of the Poisson brackets (see section 6.4.3). From this point of view the attribute *planetary* in Lagrange's planetary equations (6.99) can hardly be justified.

6.4.2 Lagrange's Equation for the Semi-major Axis $\boldsymbol{a}$

The Gaussian perturbation equation for the semi-major axis a was derived as a first example to introduce the problems. The same can be done for the Lagrangian version of the perturbation equations: The total time derivative of the perturbation function R may be computed as

$$\frac{dR}{dt} = \nabla_r R \cdot \dot{\boldsymbol{r}} + \frac{\partial R}{\partial t} = -\frac{\partial R}{\partial T_0} + \frac{\partial R}{\partial t} \, , \tag{6.101}$$

where use was made of eqn. (6.30) to represent the perturbation term.

Note, that the scalar product $\nabla_r R \cdot \dot{\boldsymbol{r}}$ may be interpreted as the time derivative of the perturbation function $R(\boldsymbol{r}(t), t)$, when ignoring the explicit time dependence of this function. In this case, the time-dependence of vector $\boldsymbol{r}(t)$ is eventually governed by the osculating mean anomaly $\sigma(t) = n\,(t - T_0)\,$, showing that the derivative w.r.t. time t and T_0 are equal, but of opposite sign. This explains the second of eqns. (6.101). Therefore, eqns. (6.101) lead directly to the final form of the perturbation equation for the semi-major axis a in the Lagrangian form:

$$\dot{a} = \frac{2\,a^2}{\mu}\,\dot{\boldsymbol{r}} \cdot \delta\boldsymbol{f} = \frac{2\,a^2}{\mu}\,\dot{\boldsymbol{r}} \cdot \nabla_r R = -\,\frac{2\,a^2}{\mu}\,\frac{\partial R}{\partial T_0} \; . \tag{6.102}$$

Equation (6.102) actually is of the general form (6.99), where, obviously, four of the five Poisson brackets showing up in the equation for the semi-major axis are zero (see next section).

The above derivation of Lagrange equation is sometimes given in textbooks as an example. It is, in a way, a nice example, because the derivation and the result are rather simple. The example is, however, not really instructive, because it is not possible to derive the perturbation equations for the other five elements along similar lines. We have to find a general method valid for all elements. This is the purpose of the next section.

6.4.3 Lagrange's Planetary Equations

The general method to derive the Lagrangian perturbation equations is implicitly given by their general form (6.99): We first have to derive the position and velocity-gradients of the elements, then we have to compute the Poisson brackets. Keeping in mind that we already have derived the expressions for the velocity-gradients (summarized in eqns. (6.83)) the remaining task cannot be too difficult. It will be solved in the next paragraph.

The Position-gradients of the Osculating Elements. The elements p, i and Ω are all derived from the angular momentum vector $\boldsymbol{h} = \boldsymbol{r} \times \dot{\boldsymbol{r}}$. As the vector product is not commutative, but changes sign when the order of its factors is reversed, we may easily derive the position-gradients from the velocity-gradients by replacing $\boldsymbol{r}$ by $\dot{\boldsymbol{r}}$ and by reversing the sign in the final formula for the position-gradients (vectorial notation). The final result reads as follows (for convenience and comparison we include the velocity-gradients, as well):

$$\begin{aligned}
\nabla_v \, p &= \frac{2}{\mu} \, (\boldsymbol{h} \times \boldsymbol{r}) \quad ; \quad & \nabla_r \, p &= -\, \frac{2}{\mu} \, (\boldsymbol{h} \times \dot{\boldsymbol{r}}) \\
\nabla_v \, i &= \frac{1}{h \, \sin i} \, (\boldsymbol{r} \times \boldsymbol{x}) \; ; \quad & \nabla_r \, i &= -\, \frac{1}{h \, \sin i} \, (\dot{\boldsymbol{r}} \times \boldsymbol{x}) \\
\nabla_v \, \Omega &= \frac{\boldsymbol{r} \cdot \boldsymbol{e}_3}{h^2 \, \sin^2 i} \, \boldsymbol{h} \quad ; \quad & \nabla_r \, \Omega &= -\, \frac{\dot{\boldsymbol{r}} \cdot \boldsymbol{e}_3}{h^2 \, \sin^2 i} \, \boldsymbol{h} \; .
\end{aligned} \tag{6.103}$$

The eccentricity e and the argument of pericenter emerge from vector $\boldsymbol{q}$ (see eqns. (6.32)). Let us focus on the eccentricity first. From the identity

$$q^2 = \boldsymbol{q} \cdot \boldsymbol{q}$$

we conclude that

$$\begin{aligned}
2 \, q \, \nabla q &= 2 \, \nabla \left\{ \boldsymbol{q} \cdot \left[\left(\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} \right) \boldsymbol{r} - (\boldsymbol{r} \cdot \dot{\boldsymbol{r}}) \; \dot{\boldsymbol{r}} \right] \right\} \\
&= 2 \, \nabla \left\{ \left(\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} \right) (\boldsymbol{q} \cdot \boldsymbol{r}) - (\boldsymbol{r} \cdot \dot{\boldsymbol{r}}) \, (\boldsymbol{q} \cdot \dot{\boldsymbol{r}}) \right\} \; ,
\end{aligned} \tag{6.104}$$

where (due to the factor of 2 on the right-hand side) the vector $\boldsymbol{q}$ can be considered as a constant when taking the gradient. Equation (4.33) was used to represent vector $\boldsymbol{q}$ on the right-hand side. The above formula holds for both, the position- and the velocity-gradient. One may now easily calculate the position-gradient from the above formula. For convenience we include the velocity-gradient, as well. The result is:

$$\begin{aligned}
q \, \nabla_r q &= \frac{\mu}{r^3} \, (\boldsymbol{r} \cdot \boldsymbol{q}) \, \boldsymbol{r} \; - \; (\dot{\boldsymbol{r}} \cdot \boldsymbol{q}) \, \dot{\boldsymbol{r}} \; + \; \left(\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} \right) \boldsymbol{q} \\
q \, \nabla_v q &= -\, (\dot{\boldsymbol{r}} \cdot \boldsymbol{q}) \, \boldsymbol{r} \; + \; 2 \, (\boldsymbol{r} \cdot \boldsymbol{q}) \, \dot{\boldsymbol{r}} \; - \; (\boldsymbol{r} \cdot \dot{\boldsymbol{r}}) \, \boldsymbol{q} \; .
\end{aligned} \tag{6.105}$$

It is a straightforward task to show that the above representation for the velocity-gradient can be reduced to the one given in eqns. (6.83).

As the position-gradients of both, p and e, are available now, we may also derive the position-gradient of the semi-major axis a, at this point, using the relation

$$\nabla a = \begin{cases} \dfrac{1}{1 - e^2} \; (\nabla p \, + \, 2 \, a \, e \;\; \nabla e) \; , & e < 1 \\[2ex] \dfrac{1}{e^2 - 1} \; (\nabla p \, - \, 2 \, a \, e \; \nabla e) \; , & e > 1 \; , \end{cases} \tag{6.106}$$

which is derived in exactly the same way as eqn. (6.38) starting from the definition of conic sections. Observe that the relation holds for both, position- and velocity-gradients. The position-gradient of a may, of course, also be directly derived from the "astronomical energy theorem" (4.20), which even may be the shortest method. In any case the result is (again, we include the velocity-gradient for convenience):

$$\nabla_v a = \pm \frac{2}{n^2 \, a} \, \dot{\boldsymbol{r}} \; ; \quad \nabla_r a = \pm \frac{2 \, a^2}{r^3} \, \boldsymbol{r} \; , \tag{6.107}$$

where the positive sign holds for elliptic, the negative for hyperbolic orbits.

The (position- or velocity-)gradient of the argument of pericenter ω is obtained in analogy to eqn. (4.33) as:

$$\nabla \omega = \frac{1}{q \, \sin \omega} \left\{ \cos \omega \, \nabla q \, - \, \nabla \left(\boldsymbol{e}_\Omega \cdot \boldsymbol{q} \right) \right\} \, - \, \cos i \, \nabla \Omega \; . \tag{6.108}$$

The equation holds for both, the position- and the velocity-gradients. The gradients of $q = \mu \, e$ and Ω were already given above, which is why we only have to deal with the gradient of the scalar product $\boldsymbol{e}_\Omega \cdot \boldsymbol{q}$, where we may consider $\boldsymbol{e}_\Omega$ to be a constant (the "neglected" term is taken care of as the last term in the above equation).

Using the representation (4.33) for vector $\boldsymbol{q}$, the scalar product in question reads as

$$\boldsymbol{q} \cdot \boldsymbol{e}_\Omega = \left(\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} \right) \left(\boldsymbol{r} \cdot \boldsymbol{e}_\Omega \right) \, - \, \left(\boldsymbol{r} \cdot \dot{\boldsymbol{r}} \right) \left(\dot{\boldsymbol{r}} \cdot \boldsymbol{e}_\Omega \right) \; , \tag{6.109}$$

from where the position- and velocity-gradients are computed as

$$\begin{aligned} \nabla_r \left(\boldsymbol{q} \cdot \boldsymbol{e}_\Omega \right) &= \frac{\mu}{r^3} \, \left(\boldsymbol{r} \cdot \boldsymbol{e}_\Omega \right) \boldsymbol{r} \, - \, \left(\dot{\boldsymbol{r}} \cdot \boldsymbol{e}_\Omega \right) \dot{\boldsymbol{r}} \, + \, \left(\dot{\boldsymbol{r}}^2 - \frac{\mu}{r} \right) \boldsymbol{e}_\Omega \\ \nabla_v \left(\boldsymbol{q} \cdot \boldsymbol{e}_\Omega \right) &= - \left(\dot{\boldsymbol{r}} \cdot \boldsymbol{e}_\Omega \right) \boldsymbol{r} \, + \, 2 \left(\boldsymbol{r} \cdot \boldsymbol{e}_\Omega \right) \dot{\boldsymbol{r}} \, - \, \left(\boldsymbol{r} \cdot \dot{\boldsymbol{r}} \right) \boldsymbol{e}_\Omega \; . \end{aligned} \tag{6.110}$$

The gradients of the time T_0 of pericenter passage is derived in analogous way as the corresponding perturbation equation. One simply has to take the gradient of Kepler's equation and to replace the term ∇E by taking the gradient of eqn. (4.53). When considering hyperbolic orbits one has to start from the analogue of Kepler's equation (see Table 4.2), where the auxiliary angle F is used instead of the eccentric anomaly E. We quote the result for elliptic orbits only:

$$\begin{aligned} \nabla_r T_0 &= - \frac{\sqrt{1 - e^2}}{e \, n \, \sin v} \left\{ \cos v \, \nabla_r e - \frac{r}{a^2} \, \nabla_r a \right\} - \frac{3}{2 \, a} \left(t - T_0 \right) \nabla_r a - \frac{\sqrt{1 - e^2}}{a \, r \, n \, e \, \sin v} \, \boldsymbol{r} \\ \nabla_v T_0 &= - \frac{\sqrt{1 - e^2}}{e \, n \, \sin v} \left\{ \cos v \, \nabla_v e - \frac{r}{a^2} \, \nabla_v a \right\} - \frac{3}{2 \, a} \left(t - T_0 \right) \nabla_v a \; . \end{aligned} \tag{6.111}$$

The gradients of a should be replaced by a linear combination of the gradients of p and e using eqn. (6.106), if p instead of a is used as osculating element. Let us briefly sketch the derivation of the position-gradient of T_0. Taking the position-gradient of Kepler's equation results in:

$$\nabla_r T_0 = -\frac{3}{2a}\,(t - T_0)\,\nabla_r a + \frac{\sin E}{n}\,\nabla_r e - \frac{r}{a\,n}\,\nabla_r E \ . \tag{6.112}$$

With the exception of $\nabla_r E$ all gradients on the right-hand side of eqn. (6.112) are available. In analogy to the computation of $\dot{E}$ when deriving the Gaussian equation for T_0 we obtain the position-gradient of E using eqn. (4.53):

$$\nabla_r r = \frac{\boldsymbol{r}}{r} = \nabla_r a \; (1 - e\cos E) \; - \; a\,\cos E \; \nabla_r e \; + \; a\,e\sin E \; \nabla_r E \ . \tag{6.113}$$

Obviously, the position-gradient of r on the left-hand side of the above equation is *not* equal to the zero-vector. Using the above result in eqn. (6.112) leads to the preliminary result:

$$\begin{aligned}\nabla_r T_0 = &\left\{-\frac{3}{2\,a}\,(t - T_0) + \frac{r^2}{a^3}\,\frac{1}{n\,e\,\sin E}\right\}\nabla_r a \\ &+\frac{1}{n}\left\{\sin E - \frac{r}{a\,e}\,\cot E\right\}\,\nabla_r e \; - \; \frac{1}{a^2\,n\,e\,\sin E}\,\boldsymbol{r} \ .\end{aligned} \tag{6.114}$$

If we replace the eccentric anomaly E by the true anomaly v using eqns. (4.51) we obtain the second of eqns. (6.111).

Poisson Brackets. The results presented in the previous paragraph allow it to calculate the Poisson brackets rather easily. The result is amazingly simple: All but five of the 15 independent Poisson brackets (when using the classical six orbital elements a, e, i, Ω, ω, and T_0) are zero. One recognizes, e.g., that the four Poisson brackets of the elements i and Ω with a and e are zero – because the gradients of i and Ω stand normal on the orbital plane and the gradients of the a and e lie in this plane. This in turn implies the products of the position-gradient (velocity-gradient) of either i or Ω with the velocity-gradient (position-gradient) of either a or e to be zero. A little bit more algebraic work is involved when evaluating the remaining 11 brackets. The actual work can, however, be left as an exercise to the reader.

The final result is contained in the Table 6.1. We already proved that the Poisson brackets are anti-symmetric, which is why we only have to provide 15 of the 36 brackets. With the Poisson brackets contained in Table 6.1 and with the general formula (6.99) of Lagrange's planetary equations it is no problem to quote the explicit version of these equations in the classical orbital elements:

Table 6.1. Poisson brackets of the classical orbital elements

	$e\,]$	$i\,]$	$\Omega\,]$	$\omega\,]$	$T_0\,]$
$[\,p,$	0	0	0	$\dfrac{2\sqrt{\lvert 1-e^2\rvert}}{n\,a}$	0
$[\,a,$	0	0	0	0	$-\dfrac{2}{n^2\,a}$
$[\,e,$	–	0	0	$-\dfrac{\sqrt{\lvert 1-e^2\rvert}}{n\,a^2\,e}$	$-\dfrac{1-e^2}{n^2\,a^2\,e}$
$[\,i,$	–	–	$-\dfrac{1}{n\,a^2\,\sin i\,\sqrt{\lvert 1-e^2\rvert}}$	$\dfrac{\cot i}{n\,a^2\,\sqrt{\lvert 1-e^2\rvert}}$	0
$[\,\Omega,$	–	–	–	0	0
$[\,\omega,$	–	–	–	–	0

$$
\begin{aligned}
\dot{a} &= \mp\frac{2}{n^2\,a}\,\frac{\partial R}{\partial T_0}\\
\dot{e} &= -\frac{\sqrt{\lvert 1-e^2\rvert}}{n\,a^2\,e}\,\frac{\partial R}{\partial \omega} - \frac{1-e^2}{n^2\,a^2\,e}\,\frac{\partial R}{\partial T_0}\\
\frac{di}{dt} &= -\frac{1}{n\,a^2\,\sqrt{\lvert 1-e^2\rvert}\,\sin i}\,\frac{\partial R}{\partial \Omega} + \frac{\cot i}{n\,a^2\,\sqrt{\lvert 1-e^2\rvert}}\,\frac{\partial R}{\partial \omega}\\
\dot{\Omega} &= \frac{1}{n\,a^2\,\sqrt{\lvert 1-e^2\rvert}\,\sin i}\,\frac{\partial R}{\partial i}\\
\dot{\omega} &= \frac{\sqrt{\lvert 1-e^2\rvert}}{n\,a^2\,e}\,\frac{\partial R}{\partial e} - \frac{\cot i}{n\,a^2\,\sqrt{\lvert 1-e^2\rvert}}\,\frac{\partial R}{\partial i}\\
\dot{T}_0 &= \frac{2}{n^2\,a}\,\frac{\partial R}{\partial a} + \frac{1-e^2}{n^2\,a^2\,e}\,\frac{\partial R}{\partial e}\,.
\end{aligned}
\tag{6.115}
$$

Observe that Table 6.1 allows it to write down Lagrange's equations for the set of elements p, e, i, Ω, ω, and T_0. The result is:

$$
\begin{aligned}
\dot{p} &= \frac{2\sqrt{|1-e^2|}}{n\,a}\,\frac{\partial R}{\partial \omega}\\
\dot{e} &= -\frac{\sqrt{|1-e^2|}}{n\,a^2\,e}\,\frac{\partial R}{\partial \omega} - \frac{1-e^2}{n^2\,a^2\,e}\,\frac{\partial R}{\partial T_0}\\
\frac{di}{dt} &= -\frac{1}{n\,a^2\sqrt{|1-e^2|}\,\sin i}\,\frac{\partial R}{\partial \Omega} + \frac{\cot i}{n\,a^2\sqrt{|1-e^2|}}\,\frac{\partial R}{\partial \omega}\\
\dot{\Omega} &= \frac{1}{n\,a^2\sqrt{|1-e^2|}\,\sin i}\,\frac{\partial R}{\partial i}\\
\dot{\omega} &= -\frac{2\sqrt{|1-e^2|}}{n\,a}\,\frac{\partial R}{\partial p} + \frac{\sqrt{|1-e^2|}}{n\,a^2\,e}\,\frac{\partial R}{\partial e} - \frac{\cot i}{n\,a^2\sqrt{|1-e^2|}}\,\frac{\partial R}{\partial i}\\
\dot{T}_0 &= \frac{1-e^2}{n^2\,a^2\,e}\,\frac{\partial R}{\partial e}\ .
\end{aligned}
\tag{6.116}
$$

From the purist's point of view it would be preferable to use consequently the elements p and e (and not a and e) in the above perturbation equations. The advantage of the above formulation is their close relationship with the classical eqns. (6.115).

We have given Lagrange's planetary equations (6.115) (or (6.116)) for the orbit of *one* celestial body in a given potential, consisting of the two-body term and the perturbation function R. In this form Lagrange's planetary equations are well suited to describe the orbital motion of any celestial body for which it is possible to compose the resulting acceleration as the sum of the two-body term and a gradient of a scalar perturbation function. The equations may be used in particular to describe the orbital motion of an artificial Earth satellite or of a minor planet in conservative force fields.

The term "planetary" indicates, that Lagrange's equations were originally meant to describe the motion of the entire planetary system. Note that in this case one set of equations of type (6.115) holds for each planet $i = 1, 2, \ldots$, and that there is one planet-specific perturbation function R_i for each of the considered planets (see eqn. (3.24)):

$$
R_i = k^2 \sum_{j=1, j\neq i}^{n} m_j \left\{ \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \frac{\boldsymbol{r}_i \cdot \boldsymbol{r}_j}{r_j^3} \right\}\ . \tag{6.117}
$$

For conservative perturbation problems in Celestial Mechanics, (i.e., when the perturbation function(s) may be described by the position-gradient of a potential), Lagrange's planetary equations and the Gaussian version (6.88) of the perturbation equations are mathematically equivalent.

For theoretical work aiming at a formal solution of the equations of motion Lagrange's version of the perturbation equations is preferable, because only one scalar function (for each of the perturbed bodies) has to be dealt with, whereas the three components R', S', and W' (or any other set of three components of the perturbing acceleration) have to be considered when using the Gaussian version. Moreover, the coefficients of the partial derivatives of the perturbation function w.r.t. the orbital elements, i.e., the Poisson brackets, are first integrals of the underlying two-body problem. A similar statement does not hold for the Gaussian version of the perturbation equations. The Gaussian version (6.88) of the perturbation equations must be used, however, as soon as non-conservative perturbing accelerations occur.

6.5 First- and Higher-Order Perturbations

On the right-hand sides of the perturbation equations in the Gaussian form (6.25) and in the Lagrangian form (6.115) there are only "small" terms of the order of the perturbation terms.

It is tempting to solve the perturbation equations (6.25) or (6.115) approximatively *by using the two-body approximation* to calculate the right-hand sides of these equations. In this approximation the right-hand sides of the equations are *known functions of time.* The error committed by this procedure is "small of the second order" in the perturbations because

- the right-hand sides are small quantities, and because
- the difference vector $\Delta\boldsymbol{r}(t) = \boldsymbol{r}(t) - \boldsymbol{r}_0(t)$ between the true solution $\boldsymbol{r}(t)$ of the system (6.25) or (6.115) and its two-body approximation $\boldsymbol{r}_0(t)$ is a small quantity, as well (at least in the vicinity of the initial epoch t_0).

By replacing $\boldsymbol{r}(t)$ by $\boldsymbol{r}_0(t)$ on the right-hand sides of the equations we are thus committing an error of the second order in the perturbations.

The impact of the approximation on the mathematical structure of the problem is important: The problem of solving a system of six coupled, non-linear ordinary differential equations is reduced to the solution of six integrals, which may be solved independently. It is thus possible to study the impact of a perturbation separately for each orbital element, or only for one or several elements of special interest.

The solution of the perturbation equations using the two-body approximation on the right-hand sides of the perturbation equations is called *the first-order solution* of the perturbation equations, and the theory outlined here is called *first-order perturbation theory.*

Let us state explicitly the first-order perturbation equation and its solution for the semi-major axis in its Gaussian form using the (T', N', W')-decomposition of the perturbing acceleration. The result must be distinguished from the true result. It may be directly transcribed from eqn. (6.87):

$$
\begin{aligned}
\dot{a}^{[1]} &= \frac{2}{n_0^2\, a_0}\, |\dot{\boldsymbol{r}}_0|\, T' \\
a^{[1]}(t) &\overset{\text{def}}{=} \frac{2}{n_0^2\, a_0} \int\limits_{t_0}^{t} |\dot{\boldsymbol{r}}_0(t')|\; T'\big(\boldsymbol{r}_0(t'), \dot{\boldsymbol{r}}_0(t'), t'\big)\, dt' \ .
\end{aligned}
\tag{6.118}
$$

The notation $a^{[1]}(t)$ indicates that the equation (6.87) was only solved using first-order perturbation theory. All quantities on the right-hand side of this equation are approximated by the two-body solution of the underlying initial-value problem.

At least in the vicinity of the osculation epoch one may expect that the first-order solution $a^{[1]}(t)$ is much better than the zero-order solution $a^{[0]} \overset{\text{def}}{=} a_0$ (which simply would be the osculating semi-major axis at the epoch t_0).

One should not forget that the first-order solution of the perturbation equations is the solution of a problem differing considerably from the original one. Nevertheless, a first-order solution, in particular when limited to a time interval of, let us say, a few dozen revolutions, usually is an excellent approximation for the true solution. It is moreover an excellent tool to gain insight into the structure of a particular problem.

Having established the first-order solution for *all six* orbital elements it is in principle easy to generate a supposedly better, second-order, solution by using the first-order solution on the right-hand sides of the perturbation equations instead of the two-body solution. The principle may be generalized to any order by using the next lower-order approximation on the right-hand side of the perturbation equations. Higher-order solutions for the semi-major axis are, e.g., characterized by:

$$
\begin{aligned}
\dot{a}^{[I+1]} &= \frac{2}{n^{[I]}\, a^{[I]}}\, |\dot{\boldsymbol{r}}^{[I]}|\, T'^{[I]} \\
a^{[I+1]}(t) &\overset{\text{def}}{=} \int\limits_{t_0}^{t} \frac{2}{n^{[I]}(t')\, a^{[I]}(t')}\, |\dot{\boldsymbol{r}}^{[I]}(t')|\; T'^{[I]}\left(\boldsymbol{r}^{[I]}(t'), \dot{\boldsymbol{r}}^{[I]}(t'), t'\right)\, dt' \ .
\end{aligned}
\tag{6.119}
$$

Note, that the above integral may be solved independently from the integrals for the other five elements. The solution method of the perturbation problem is thus the same in the higher than the first order: only known functions of time are used in the integration process. One should be aware of the fact, however, that on the right-hand side of the above equation a complete solution of order I is needed in order to accomplish a correct solution of order $I+1$, even if only one of the orbital elements is of interest.

The procedure for generating first-order, then higher-order approximations of the original perturbation equations is straight forward – and usually, it works.

There are, however, risks involved in the procedure. Two of them should be mentioned:

- If a first-order solution is used over very long time periods (hundreds of revolutions), the essential assumption of first-order theory, namely that the difference between the true and the two-body solution is a “small” quantity, may become violated. This may lead to convergence problems, if the attempt is made to generate higher-order solutions.
- If solutions are sought in the tradition of general perturbation theory, it may occur that a perturbation term is exactly periodic in a first-order approximation. It may in addition happen that small divisors or even a zero-divisor occur, when taking the time derivative of this perturbation,. Such problems are in particular encountered when resonance problems are studied.

If first- and higher-order solutions are established using numerical methods, such problems usually do not occur. In order to avoid problems of this kind, one has to brake up a long time-interval into shorter partial intervals and apply first- and higher-order approximation within these shorter intervals. Numerical methods for solving the transformed equations of motion are, however, not yet well established. They should be considered as promising and very powerful tools.

It is refreshing to read Taff’s treatise on perturbation theory, in particular the section “the misapplications of perturbative theory”, which is concluded by the editorial starting with the statement (see L. G. Taff, [118]): *” . . . beyond first-order theory I know of no useful result from perturbation theory in Celestial Mechanics because all of the higher-order results have no firm mathematical basis. Frequently the second approximation produces nonsensical results . . .* ” We agree with this assessment, where analytical theories (method of general perturbations) are concerned – and such theories were meant in the treatise [118]. We are, however, much more optimistical concerning the application of the perturbation approach in the field of special perturbation theory (i.e., when solving the perturbation equations in the elements using the technique of numerical integration).

6.6 Development of the Perturbation Function

In general perturbation theory an “analytical” solution of the perturbation equations is sought. This implies that the perturbation function has to be developed into a series, the terms of which may be integrated in closed form (“analytically”). General perturbation methods are very well suited to gain a quick overview of a particular perturbation function. In such analyses first-order perturbation theory is sufficient and usually it is allowed to simplify the problem considerably.

If higher accuracy is required, the development of the perturbation function may frustratingly complicated, and it is, by definition, problem-dependent. This section will be concluded by a simple (simplified) example to demonstrate the technical difficulties involved in this step. Other examples may be found in subsequent chapters.

Until quite recently general perturbation theory was the only tool available for practical problems requiring highest accuracy in Celestial Mechanics. The striking example is Newcomb's theory of the planetary (and natural satellite) motion, which provided the firm fundament for astronomical almanacs well into the second half of the 20th century. Only with the advent of fast computers (and this time we do not speak of human beings) it was possible to circumvent and, what was at least as important, to check analytical theories of motion in Celestial Mechanics by numerical solutions.

Subsequently, we will use general perturbation theory only to generate simple approximations (and apply them only to short time intervals). The following example shall illustrate the use and limitations of first-order perturbation theory.

6.6.1 General Perturbation Theory Applied to Planetary Motion

Let us study the orbital motion of two planets with small eccentricities in one and the same orbital plane around the Sun. Formally, the distinction is made between the perturbing and the perturbed planet. *No* index will be used for the perturbed body and the index is p reserved for the perturbing body. According to eqn. (6.117) the perturbing function (for the perturbed body) may be written as:

$$R = k^2 \, m_p \left\{ \frac{1}{|\boldsymbol{r} - \boldsymbol{r}_p|} - \frac{\boldsymbol{r} \cdot \boldsymbol{r}_p}{\boldsymbol{r}_p^3} \right\} \ . \qquad (6.120)$$

As we are dealing with a planar problem, we may disregard the perturbations in the inclination i and the longitude of the node Ω and introduce instead the longitude of the perihelion, the sum of the longitude of the node and the argument of perihelion:

$$\begin{aligned} \tilde{\omega} &\overset{\text{def}}{=} \Omega + \omega \\ \tilde{\omega}_p &\overset{\text{def}}{=} \Omega_p + \omega_p \ . \end{aligned} \qquad (6.121)$$

Neglecting the terms of higher than first order in the eccentricities e and e_p the following approximations may be used for quantities related to the perturbed body:

$$\begin{aligned}
\sigma &= n\,(t-T_0)\\
E &\approx \sigma + e\sin\sigma\\
\cos E &= \cos\sigma - \tfrac{1}{2}\,e\,(1-\cos 2\sigma)\\
\sin E &= \sin\sigma + \tfrac{1}{2}\,e\,\sin 2\sigma\\
r &= |\boldsymbol{r}| = a\,(1-e\cos E)\\
&\approx a\,(1-e\cos\sigma)\\
r^{-m} &\approx a^{-m}\,(1+m\,e\,\cos\sigma)\ , \quad m=1,2,\ldots\\
r_1 &\approx a\big(\cos(E+\tilde{\omega}) - e\,\cos\tilde{\omega}\big)\\
&\approx a\left(\cos(\sigma+\tilde{\omega}) + \frac{e}{2}\,\cos(2\,\sigma+\tilde{\omega})\right)\\
r_2 &\approx a\big(\sin(E+\tilde{\omega}) - e\,\sin\tilde{\omega}\big)\\
&\approx a\left(\sin(\sigma+\tilde{\omega}) + \frac{e}{2}\,\sin(2\,\sigma+\tilde{\omega})\right)\ ,
\end{aligned} \tag{6.122}$$

where r_1 and r_2 are two components of the position vector of the perturbed body in the orbital plane, where the first coordinate axis pointing to the perihelion. Analogous approximations are obtained for the perturbing body.

With these approximations we are now in a position to develop the perturbing function into a series with integrable terms. We may assume that

$$|\boldsymbol{r}| > |\boldsymbol{r}_p|\ , \tag{6.123}$$

which allows it to develop the first (and crucial) term of the perturbation function (6.120) into series of Legendre polynomials (see eqn. (3.101))

$$\frac{1}{|\boldsymbol{r}-\boldsymbol{r}_p|} = \frac{1}{|\boldsymbol{r}|}\sum_{i=0}^{\infty}\left(\frac{|\boldsymbol{r}_p|}{|\boldsymbol{r}|}\right)^i P_i\left(\frac{\boldsymbol{r}\cdot\boldsymbol{r}_p}{|\boldsymbol{r}|\;|\boldsymbol{r}_p|}\right)\ , \tag{6.124}$$

where the functions $P_i(x)$ are the Legendre polynomials of degree i. Using the approximations (6.122) the argument of the Legendre polynomials may be written as:

$$\begin{aligned}
\frac{\boldsymbol{r}\cdot\boldsymbol{r}_p}{|\boldsymbol{r}|\;|\boldsymbol{r}_p|} &= \big\{\cos(\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p) + \tfrac{1}{2}\,e\,\cos(2\,\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p)\\
&\quad + \tfrac{1}{2}\,e_p\,\cos(2\,\sigma_p-\sigma+\tilde{\omega}_p-\tilde{\omega})\big\}\big\{1+e\,\cos\sigma+e_p\,\cos\sigma_p\big\}\\
&= \cos(\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p)\\
&\quad + \tfrac{1}{2}\,e\,\big\{2\,\cos(2\,\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p) + \cos(\sigma_p-\tilde{\omega}+\tilde{\omega}_p)\big\}\\
&\quad + \tfrac{1}{2}\,e_p\,\big\{2\,\cos(2\sigma_p-\sigma+\tilde{\omega}_p-\tilde{\omega}) + \cos(\sigma+\tilde{\omega}-\tilde{\omega}_p)\big\}\ .
\end{aligned} \tag{6.125}$$

In the development (6.124) we have to raise this argument up to power n, resulting in expressions of the kind

$$\left(\frac{\boldsymbol{r}\cdot\boldsymbol{r}_p}{|\boldsymbol{r}|\,|\boldsymbol{r}_p|}\right)^n = \cos^n(\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p) \;+\; n\,\cos^{n-1}(\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p)$$
$$\cdot\Big[+\tfrac{1}{2}\,e\,\big\{2\,\cos(2\,\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p)+\cos(\sigma_p-\tilde{\omega}+\tilde{\omega}_p)\big\}$$
$$+\tfrac{1}{2}\,e_p\,\big\{2\cos(2\sigma_p-\sigma+\tilde{\omega}_p-\tilde{\omega})+\cos(\sigma+\tilde{\omega}-\tilde{\omega}_p)\big\}\Big]\;. \tag{6.126}$$

Using the standard trigonometric relations one may replace the nth power of a cosine-function by a linear combination of cosine-functions of multiples of its argument. It is therefore possible to write the perturbation function as a trigonometric series in the multiples of the angles $i\,\sigma-k\,\sigma_p+j\,(\tilde{\omega}-\tilde{\omega}_p)$. The structure of the resulting series development is:

$$\begin{aligned}\tilde{R} &= k^2\,m_p\left\{\frac{1}{|\boldsymbol{r}-\boldsymbol{r}_p|}-\frac{\boldsymbol{r}\cdot\boldsymbol{r}_p}{\boldsymbol{r}_p^3}\right\}\\ &=\sum_{j=0}^{+\infty}\alpha_j(a,a_p)\,\cos\big[j\,(\sigma-\sigma_p+\tilde{\omega}-\tilde{\omega}_p)\big]\\ &\quad+e\sum_j\sum_k\sum_l\beta_{jkl}(a,a_p)\,\cos\big[j\,\sigma-k\,\sigma_p+l\,(\tilde{\omega}-\tilde{\omega}_p)\big]\\ &\quad+e_p\sum_j\sum_k\sum_l\gamma_{jkl}(a,a_p)\,\cos\big[j\,\sigma-k\,\sigma_p+l\,(\tilde{\omega}-\tilde{\omega}_p)\big]\;.\end{aligned} \tag{6.127}$$

The sum limits were not specified in the above expression. It is important, however, that in the terms proportional to e and e_p we do not only have terms with arguments proportional to $l\,(\sigma-\sigma_p)+\ldots$.

In first-order theory, all of the above angular arguments *except* the two mean anomalies σ and σ_p do *not* depend on time.

We are now ready to use the development (6.127) in Lagrange's planetary equations (6.115), where the partial derivatives of the development (6.127) w.r.t. the orbital elements have to be calculated.

The development (6.127) shows that the dependence on the eccentricity e is only contained in the coefficients of the development, whereas the dependence on T_0 and $\tilde{\omega}$ resides uniquely in the arguments of the cos-functions. With the exception of the semi-major axis, which occurs in the coefficients and implicitly in the anomalies (the mean motion n is a function of a), the dependence on a particular orbital element is contained either in the coefficients or in the arguments of the cos-series, but not in both.

As the arguments of the trigonometric series are linear functions of the mean anomalies, and as the mean anomalies in turn are linear functions of time, the differential equations resulting after replacing the perturbation function on the right-hand sides of Lagrange's planetary equations (6.115) may be integrated formally. The individual terms give raise to

- *periodic perturbations* with periods $P = \frac{2\pi}{j\,n - k\,n_p}$, or
- *secular perturbations*, growing linearly with time t.

The basic period of the system occurring in the term proportional to $e^0 e_p^0$ is the synodic revolution period of the two planets

$$P = \frac{2\,\pi}{|\,n - n_p\,|} \ . \tag{6.128}$$

Perturbations with periods of the order of P are called *short-periodic.*

Secular perturbations only occur, if there are terms *not* depending on the mean anomalies in the integrands. Such terms cannot show up, if the partial derivative of expression (6.127) is taken w.r.t. T_0, because all resulting terms will contain $j\sigma$ with $\sigma \neq 0, \ j \neq 0$ in the argument. This is why secular perturbations cannot occur in the semi-major axis a in first-order theory. This might be viewed as an argument for the stability of the planetary system. Unfortunately the argument is not valid because secular terms might show up in higher-order solutions.

Because we have to take the partial derivative of the perturbing function w.r.t. the eccentricity e in the equation for the perihelion (and because the e-dependence is contained in the coefficients of the development), secular terms may occur when solving the equation for the perihelion. The examples in Chapter II- 4 will show that secular terms actually do show up in the perihelion and in the longitude of the ascending node. The revolution periods of these angles are very long compared to the basic period (the synodic revolution period P) of the system. In higher-order approximations one would thus expect perturbations with periods related to the revolution period of both, the longitude of the pericenter and the longitude of the ascending node.

Long-period perturbations may, however, already show up in first-order solutions, if

$$|\,j\,n \, - \, k\,n_p\,| \approx 0 \ . \tag{6.129}$$

The relation (6.129) holds, if the two mean motions are nearly *commensurable.* The amplitudes of the corresponding perturbation terms may be rather big, despite the fact that they are proportional to e and e_p, because they are amplified (after the integration) by the factor $|\,j\,n \, - \, k\,n_p\,|^{-1}$.

Through this mechanism long-period perturbations with considerable amplitudes may show up even in first-order theory, provided the revolution periods are nearly commensurable. The *great inequality* of the planets Jupiter and Saturn of about 900 years is an example for such a mechanism (see also Chapter II- 4).

If the resonance condition (6.129) holds precisely (this may occur when solving a perturbation problem with analytical means), the method, when blindly used, fails due to the occurrence of zero-divisors. The problem can of course removed by considering the associated sin- or cos-term as constant, e.g., by the approximation

$$\cos(j\,\sigma - k\,\sigma_p + \alpha) \approx \cos\left(k\,n_p\,(T_{01} - T_0) + \alpha\right) . \tag{6.130}$$

A term of this kind is capable of producing seemingly "secular" perturbations in all orbital elements, a result which may become nonsensical if used over very long time intervals. The example illustrates the difficulties and problems that may occur in the framework of general perturbation theory.

If more than one perturbing planet are acting on the perturbed body, the perturbations may be calculated separately in first-order theory and the combined effect is obtained by superposition. This makes first-order perturbation theory a very powerful and efficient tool. The theory, enhanced by selected higher-order terms, was used for the production of the ephemerides of our almanacs till the second half of the 20th century.

The above example was a gross simplification of the analytical developments actually performed for applications in the planetary system. When allowing for moderate eccentricities and inclinations between the orbital planes, the perturbation function of two bodies moving around the Sun may be written as (see, e.g., [94]):

$$R = \sum P \cos Q \, + \sum P' \cos Q' \ , \tag{6.131}$$

where the coefficients P and P' are functions of the elements a, a_p, e, e_p and i, i_p. In addition to the first-order terms in the eccentricities, higher-order terms have to be taken into account, as well. Moreover, the inclination between the orbital planes generates additional terms proportional to $\sin^k i$ and $\sin^l i_p$. The argument Q is time-independent in first-order theory, whereas Q' contains the time-dependence:

$$\begin{aligned} Q \ &= j\,\Omega \,+\, j_p\,\Omega_p \,+\, k\,\tilde{\omega} \,+\, k_p\,\tilde{\omega}_p \\ Q' &= l\,n\,(t - T_0) \,+\, l_p\,n_p\,(t - T_{10}) \,+\, j\,\Omega \,+\, j_p\,\Omega_p \,+\, k\,\tilde{\omega} \,+\, k_p\,\tilde{\omega}_p \ . \end{aligned} \tag{6.132}$$

Equations (6.131) are somewhat more general than the development (6.127). The elegant shape of eqns. (6.131) should not hide the fact that their explicit versions are rather complicated.

6.7 Perturbation Equation for the Mean Anomaly $\sigma(t)$

The Gaussian and Lagrangian perturbation equations (6.88) and (6.115) were set up for the osculating orbital elements a, e, i, Ω, and T_0 (or $\sigma_0(=\sigma(t_0))$).

All osculating elements are first integrals of the two-body equations of motion (4.1).

All of the Gaussian perturbation equations (6.88) except the one for the equation for the time of pericenter passage (or, alternatively, the equation for the mean anomaly σ_0) are of a similar and simple structure: The right-hand sides are linear combinations of the perturbing accelerations, and the coefficients of the linear combinations are to the first-order (when using the two-body approximation on the right-hand side) periodic functions of time (with the revolution period of the perturbed body as period). The result therefore is "well-behaved" – provided the perturbing accelerations are "reasonable", as well: One may expect periodic functions (with periods given by the perturbed body and the periods contained in the perturbing accelerations) and possibly linear functions of time.

In the equation for T_0 or σ_0 there is, however, a term proportional to the time interval $t - T_0$ or $t - t_0$. For integrations over long time intervals, this term is going to dominate all other terms, and it must generate periodic terms of a linearly growing amplitude. Figure 6.2, showing the the initial mean anomaly σ_0 (referring to January 1, 2000) as a function of time over 2000 years, illustrates the effect. The Figure is based on the osculating elements obtained by numerically integrating the outer planetary system. Whereas the amplitudes of all other perturbations are small, we observe terms with linearly growing amplitudes, which are already of the order of 30° after 2000 years. The result is not easy to interpret – as a matter of fact it is a kind of an artefact (see subsequent discussion). This is why the development of σ_0 or

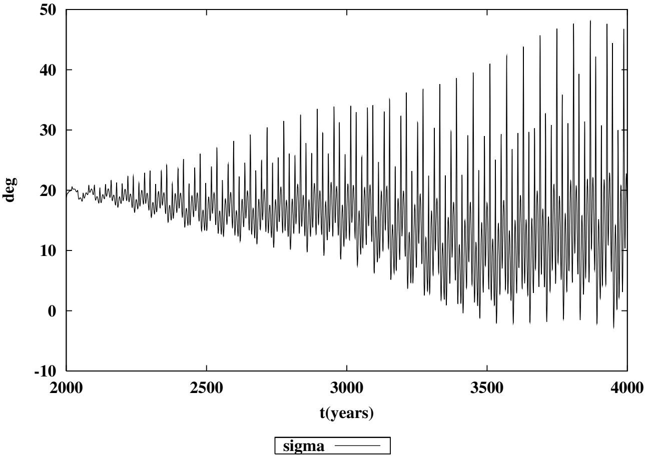


Fig. 6.2. $\sigma_0 = \sigma(2000.0)$ of Jupiter over a time interval of 2000 years

of T_0 usually is not discussed in textbooks, a tradition which will be followed here.

When inspecting Lagrange's planetary equations, one might first think that a similar problem does not show up in eqns. (6.115). This is not the case, however: the equation for T_0 contains the partial derivative of the perturbation function R w.r.t. the perturbed body's semi-major axis. As this perturbation function does contain the radius vector $\boldsymbol{r}$ of the perturbed body (or functions thereof), it will also depend on the true anomaly v, which in turn depends on a, e, and T_0 via the eccentric and the mean anomaly $\sigma(t) = n\,(t - T_0)$ (see section 5.3). This dependence shows that the term $\frac{\partial R}{\partial a}$ also must contain terms proportional to $t - T_0$, as well.

The situation can be improved because the elements σ_0 or T_0 are only needed to calculate the mean anomaly $\sigma(t)$ for the time argument t considered. Having established the perturbation equations for all six osculating elements (referring to t_0), it is of course also possible to derive a differential equation for $\sigma \overset{\text{def}}{=} \sigma(t)$. The mean anomaly $\sigma(t)$ at time t is a function of a, T_0, and t:

$$\sigma = n\,(t - T_0) = \sqrt{\frac{\mu}{a^3}}\,(t - T_0)\ .$$

Taking the time-derivative of this equation we obtain:

$$\dot{\sigma} = n\ -\ \frac{3\,n}{2\,a}\,(t - T_0)\,\dot{a}\ -\ n\,\dot{T}_0\ . \tag{6.133}$$

Using eqns. (6.88) one easily verifies that the differential equation for σ does *not* contains terms proportional to $(t - T_0)$, but only the constant term n (the two-body term) and small perturbation terms:

$$\dot{\sigma} = n\ +\ \frac{1 - e^2}{n\,a\,e}\left\{\left(\cos v - 2\,e\,\frac{r}{p}\right)R' - \left(1 + \frac{r}{p}\right)\sin v\,S'\right\}\ . \tag{6.134}$$

Equations (6.134) are easier to handle, independently of whether general or special perturbation methods are applied. If necessary, the osculating element T_0 may be reconstructed after having solved the Gaussian perturbation equations (6.88) for a, e, i, Ω, ω, and $\sigma(t)$ using the definition (6.7) of the mean anomaly.

Lagrange's planetary equations may be modified following the same pattern as in the case of the Gaussian equations. The differential equation for σ is again given by eqn. (6.133), but we have to replace $\dot{T}_0$ and $\dot{a}$ using eqns. (6.115):

$$\dot{\sigma} = n\ -\ \frac{3\,n}{2\,a}\,(t - T_0)\,\dot{a}\ -\ n\left[\frac{2\,a^2}{\mu}\,\frac{\partial R}{\partial a} + \frac{L^2}{\mu^2\,e}\,\frac{\partial R}{\partial e}\right]\ . \tag{6.135}$$

Now, the partial derivative w.r.t. the semi-major axis a may be split up into a derivative $\{R\}_a$, where the dependence of the true anomaly v on a is ignored, and a contribution taking into account exactly this term (obviously it is only the latter term which contains the terms proportional to the time argument $t - T_0$):

$$\begin{aligned} \frac{\partial R}{\partial a} &= \{R\}_a + \frac{\partial R}{\partial v} \frac{\partial v}{\partial E} \frac{\partial E}{\partial a} \\ &= \{R\}_a + \frac{3}{2\,a}\,(t - T_0)\,\frac{\partial R}{\partial v} \frac{\partial v}{\partial E} \frac{\partial E}{\partial T_0} \\ &= \{R\}_a + \frac{3}{2\,a}\,(t - T_0)\,\frac{\partial R}{\partial T_0} \\ &= \{R\}_a - \frac{3}{4}\,n^2\,(t - T_0)\,\dot{a}\ . \end{aligned} \tag{6.136}$$

Replacing in eqn. (6.135) the partial derivative of R w.r.t. a by the above expression gives the following differential equation for the mean anomaly $\sigma(t)$:

$$\dot{\sigma} = n - \frac{2}{n\,a}\,\{R\}_a - \frac{1 - e^2}{n\,a^2\,e}\,\frac{\partial R}{\partial e}\ . \tag{6.137}$$

Equation (6.137) is the equivalent to the perturbation equation for σ in the Gaussian formulation (6.134). Both versions do *not* contain terms proportional to the time t and are thus much more convenient to use than the equations for σ_0 or for T_0. Keep in mind that the symbol $\{R\}_a$ stands for the partial derivative of $R(a, e, v)$ w.r.t. a, disregarding the fact that the true anomaly v *also* depends on the semi-major axis a.

Figure 6.3 proves that the alternative "orbital element" $\sigma = \sigma(t)$ is much better behaved than σ_0. σ only shows a linear trend and periodic variations with a constant amplitude – exactly as all the other osculating elements. Observe, that a linear trend of $n_0\,(t - t_0)$, where n_0 is the osculating mean motion at $t = t_0$, was removed in Figure 6.3 in order to make the periodic variations visible.

A Side-Issue: The problem of "time t outside the trigonometric arguments" is well known and may be dealt with in different ways. Brouwer and Clemence [27], Roy [94], and most of the other "modern" authors introduce a "new" element which does not contain the terms mentioned. They then derive a second-order differential equation for this auxiliary element. Kaula [62] even derives the Lagrangian perturbation equations using the set of parameters a, e, i, Ω, ω, and $\sigma(t)$ throughout the development. His result is, as a matter of fact, equivalent to the equations presented here, because $\frac{\partial R(t)}{\partial T_0} = -\,n\,\frac{\partial R(t)}{\partial \sigma}$. We believe that the equations given above for the five orbital elements a, e, i, Ω, ω and for the mean anomaly σ are very easy to understand and (what may be even more important) to use.

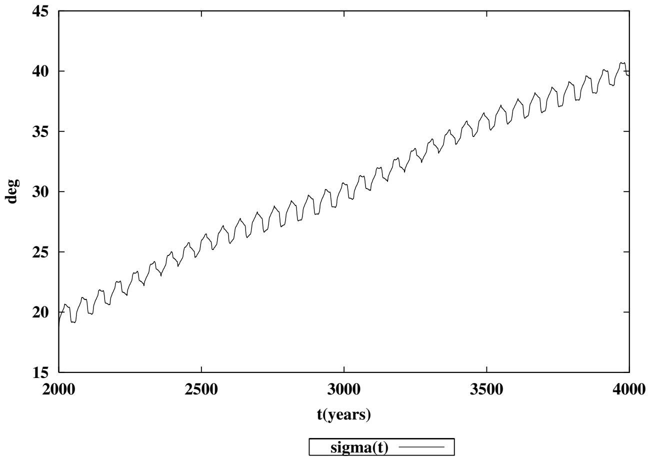


Fig. 6.3. De-trended mean anomaly $\sigma(t) - n_0(t - t_0)$ of Jupiter over a time interval of 2000 years

7. Numerical Solution of Ordinary Differential Equations: Principles and Concepts

7.1 Introduction

The three different types of equations of motion set up in Chapter 3 and the associated variational equations discussed in Chapter 5 are the central theme of this book. In Chapter 6 it was shown that the equations of motion, when using the osculating orbital elements instead of the celestial bodies' coordinates and velocities as dependent variables, may be represented approximately as integrals of known functions of time. It seems therefore wise to put the discussion of solution methods of differential equation systems on a general basis in order to cope with quadrature, with linear, as well as non-linear differential equation systems of any order.

In Celestial Mechanics the equations of motions are in general so complex that only numerical methods promise efficient, yet accurate solutions. It is therefore appropriate to develop the key algorithms of numerical analysis related to the solution of differential equation systems and to numerical quadrature. We cannot strive for completeness, but we include those solution methods which are of greatest importance in Celestial Mechanics.

The numerical solution of ordinary differential equation systems should not be understood (and taught) as a "catalogue of recipes". We will restrict the discussion to few fundamental principles and concepts of numerical analysis.

In pure mathematics manifolds of solutions of differential equation systems may be discussed. In numerical analysis the focus is on particular solutions of ordinary differential equation systems. A particular solution is defined by the differential equation system and additional information, usually the initial values of the solution at one particular value t_0 of the independent argument t. We will also address what might be called a local boundary value problem (see sections 7.2 and 7.5). The initial and the local boundary value problems will be defined more precisely in the next section.

When dealing with ordinary differential equation systems the distinction is made between one independent and one or more dependent arguments. A particular solution represents all dependent arguments as functions of the independent argument. Dependent and independent arguments have different meanings in different applications. In an attempt to keep the language simple, the symbol t will be used for the independent argument and it is identified with time. This convention allows it to speak, e.g., of an initial epoch, of boundary epochs, etc.

Solving an initial or a boundary value problem should be viewed as a special task of approximation theory, where an approximating function for the true solution is sought. From now on it shall be called the numerical solution of the underlying mathematical problem. The numerical solution might be a truncated Taylor series (named after Brook Taylor (1685–1731)), a series of trigonometric functions, and so on. Special problems might favor special functions. Subsequently, we are only interested in algorithms with the potential to solve "all possible" ordinary differential equation systems. When solving a differential equation system with numerical methods we will observe the following general guidelines:

- The numerical solution represents each component of the solution as a linear combination of given base functions. We will use polynomials for this purpose.
- The numerical solution is generated independently of specific user requirements (in the sense that the solution is needed at such and such epochs), except that the approximating function must cover the entire time interval of interest.
- After the actual solution step, the numerical solution may be evaluated at any required epochs, its time derivative(s) may be taken at any epoch, the numerical solution may be integrated over any time interval covered by the solution.

This *purist's understanding* of numerically solving an initial or boundary value problem is not common – and it rules out a number of well known and commonly used methods. Such alternative algorithms will only be included for comparison purposes in section 7.4.

Formally, the numerical solution of a differential equation system may be written as the sum of the true solution and of an error function. Good integration methods should have the capability to assess and/or control the errors of the numerical solution. The appropriate treatment of errors and their accumulation is probably the most complicated aspect of numerical integration methods.

Numerical integration is dealt with in six sections (not counting this introductory section). The general mathematical structure of the problems is discussed in section 7.2. The Euler method, so to speak the "mother of all integration methods", is reviewed in section 7.3. This review is mandatory, because modern methods share many – as a matter of fact most – properties with Euler's algorithm. Section 7.4 gives an overview of important and powerful integration procedures in use today for the numerical solution of ordinary differential equation systems. Section 7.5, dealing with the collocation methods, is the core of our treatment. Collocation methods may be viewed as the logical successors of the Euler method, sharing all properties with the Euler method but being orders of magnitude more efficient. The well-known multistep methods are shown to be special cases of collocation procedures in this section. In section 7.6 collocation algorithms are applied to linear differential equation systems and to definite integrals, i.e., to numerical quadrature. The famous Gaussian quadrature formulae are shown to be special cases of collocation methods, as well. The chapter is concluded with a discussion of the error propagation in section 7.7, where the emphasis is put on problems of Celestial Mechanics.

A treatment of numerical analysis without including at least a few key examples would be like a soup without salt. The illustrations in this Chapter are based on the programs NUMINT, LINEAR, and PLASYS, which are included in the program package and documented in Chapters II- 6 and II- 10 of Part III. Program NUMINT is particularly well suited to compare integration algorithms. Two basic problems of Celestial Mechanics, namely the motion of an artificial Earth satellite in the gravity field of the oblate Earth, and the motion of a minor planet in the gravity field of Sun and Jupiter, may be solved using a variety of different methods. The program LINEAR is used to solve selected linear differential equation systems and to evaluate definite integrals. Program PLASYS is used to solve the planetary N-body problem.

7.2 Mathematical Structure

The equations of orbital and rotational motion derived in Chapter 3 are special cases of the following explicit system of ordinary differential equation systems of order n :

$$\boldsymbol{y}^{(n)} = \boldsymbol{f}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \ , \tag{7.1}$$

where

$n > 0$ is the order of the differential equation system,

$d > 0$ is the dimension of the system of equations, implying that

$\boldsymbol{y} = \boldsymbol{y}(t)$ is the column array of d dependent variables $y_j(t)$, $j = 1, 2, \ldots, d$, i.e., the array of unknown functions of time,

t is the time, the independent variable of the system of differential equations,

$\dot{\boldsymbol{y}} = \dot{\boldsymbol{y}}(t)$ is the first derivative of $\boldsymbol{y}(t)$ w.r.t. t,

$\ddot{\boldsymbol{y}} = \ddot{\boldsymbol{y}}(t)$ is the second derivative of $\boldsymbol{y}(t)$ w.r.t. t,

$\boldsymbol{y}^{(i)} = \boldsymbol{y}^{(i)}(t),\ i = 1, 2, \ldots$, is the i-th derivative of $\boldsymbol{y}(t)$ w.r.t. t, and

$\boldsymbol{f} = \boldsymbol{f}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big)$ is the right-hand side of the system of differential equations.

Loosely speaking, $\boldsymbol{y}(t)$ is also called the solution vector. Note, however, that the dimension d of the system has nothing to do with the dimension of the vector space of the problem addressed. In the equations of motion of the planetary system with N planets and/or planetoids and comets, etc., the dimension of the differential equation system would be $d = 3\,N$. For our applications only equations of orders $n = 1$ and $n = 2$ are actually needed.

Equation (7.1) is not the most general formulation for a differential equation of order n. A more general formulation would be:

$$\tilde{\boldsymbol{f}}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}, \boldsymbol{y}^{(n)}\big) = \boldsymbol{0} \ . \tag{7.2}$$

Quite a few important equations of mathematical physics are of the general form (7.2), which might be called the implicit formulation of a differential equation system. All systems of type (7.1) may be written in the form (7.2), but not all systems of type (7.2) may be brought (easily) into the form (7.1). Only explicit systems of type (7.1) are considered in this book.

The system (7.1) of order n and dimension d might be written in component form

$$\begin{pmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \cdots \\ \cdots \\ y_d^{(n)} \end{pmatrix} = \begin{pmatrix} f_1\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \\ f_2\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \\ \cdots \\ \cdots \\ f_d\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \end{pmatrix} = \begin{pmatrix} f_1(t) \\ f_2(t) \\ \cdots \\ \cdots \\ f_d(t) \end{pmatrix} \ . \tag{7.3}$$

Wherever possible the matrix notation (7.1) will be given the preference over the component notation (7.3).

In general, all components of $\boldsymbol{y}(t)$ and of its first $n-1$ time derivatives occur on the right-hand sides of eqns. (7.1). In this case the system is called a *coupled* system of equations. If it is possible to split the system (7.3) into two subsystems, where, within (at least) one subsystem only the components referring to that subsystem occur on the right-hand sides, the system (7.3) is separable, and both subsystems can be solved separately. Observe that the system containing only the components of this system on the right-hand side has to be solved first.

A planetary system with a mixture of finite point masses and bodies of negligible masses is described by a separable system. Obviously, the non-zero masses give rise to a coupled system of type (3.18), which does, however, not depend on the motion of the bodies of negligible mass. This subsystem describing the motion of the finite masses must be solved first. Subsequently, we may solve – one by one – the equations of motion (3.21) for the bodies of negligible mass. Obviously, the coordinates of the particular body of negligible mass and the coordinates of all bodies with finite masses (which are known functions of time after the solution of eqns. (3.18)) occur on the right-hand side of the differential equation for the particular body.

In Chapter 5 we have seen that the variational equations associated with the equations of motion are linear. Linear systems may be written in the form

$$\boldsymbol{y}^{(n)} = \sum_{i=0}^{n-1} \mathbf{A}_i(t)\, \boldsymbol{y}^{(i)} \,+\, \boldsymbol{b}(t) \;, \tag{7.4}$$

where $\mathbf{A}_i(t)$ are square matrices of dimension d with elements $A_{i_{jk}} = A_{i_{jk}}(t)$, $j, k = 1, 2, \ldots, d$, known as functions of time t, and where $\boldsymbol{b}(t)$ is a column array, consisting of d known functions $b_j(t)$, $j = 1, 2, \ldots, d$, of time.

The linearity of the differential equation systems may, but need not, be exploited by numerical algorithms. This aspect will be further pursued in section 7.6.

When numerically solving the system (7.1) we need additional information to identify one particular solution, e.g., by defining an *initial value problem*, where, at an initial epoch t_0, the solution vector and its first $n-1$ derivatives w.r.t. time t are specified. The initial value problem reads as:

$$\begin{aligned} \boldsymbol{y}^{(n)} \quad &= \boldsymbol{f}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \\ \boldsymbol{y}^{(i)}(t_0) &\overset{\text{def}}{=} \boldsymbol{y}_0^{(i)} \;, \quad i = 0, 1, \ldots, n-1 \;. \end{aligned} \tag{7.5}$$

The solution vector and its first $n-1$ time derivatives at a particular time t also are referred to as the *state vector* of the system at time t. Consequently the initial values $\boldsymbol{y}_0^{(i)}$, $i = 0, 1, \ldots, n-1$, are referred to as the *initial state vector* at time t_0.

Equations (7.5) consist of the differential equation system, which should hold for all time arguments t (it is the system of eqns. (7.1)), and the initial values of the solution array and its first $n-1$ derivatives at time t_0. One can prove that the initial value problem (7.5) has exactly one solution, if the function $\boldsymbol{f}$ meets certain requirements (key word: Lipschitz conditions). We do not review the existence and uniqueness theorems and proofs associated with the initial problem (7.5), a topic, which is covered in many mathematical textbooks.

Particular solutions may be specified in many alternative ways. If the condition equations referring to one initial epoch t_0 are replaced by equations referring to several epochs, one generally speaks of a *boundary value problem*, which might, e.g., be formulated as follows:

$$\begin{aligned} \boldsymbol{y}^{(n)} &= \boldsymbol{f}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \\ \boldsymbol{y}(t_i') &\stackrel{\text{def}}{=} \boldsymbol{y}_i \ , \quad i = 1, 2, \ldots, n, \quad t_i' \neq t_k' \ \text{ for } \ i \neq k \ . \end{aligned} \tag{7.6}$$

In the above example the time derivatives of the solution vector at time t_0 were replaced by the solution vector at $n-1$ different epochs. It is a non-trivial task to decide whether the problem (7.6) has a unique solution. (As a matter of fact boundary value problems in Celestial Mechanics generally do not have unique solutions, but usually all but one solution may be ruled out in practice.) Two special cases are important: (a) if an approximate solution of the boundary problem is available, a solution may be found by linearization; (b) if the boundary epochs t_i', $i = 1, 2, \ldots n$, are close together (within the convergence radius of a Taylor series expansion of the solution with suitable origin), it is possible to find a numerical solution by a special technique (see section 7.5). Problems of type (b) may be called *local boundary value problems.*

More general boundary value problems than that represented by eqns. (7.6) occur in practice. One might, e.g., wish to provide the value for $\boldsymbol{y}(t)$ at time t_1', its first derivative at time t_2', etc. One might even think of specifying different time derivatives for different components at one and the same time. It would not be too difficult to develop algorithms for such general situations. In order to focus on problems actually occurring in Celestial Mechanics, we refrain from studying such cases and uniquely deal with the initial value problem (7.5) and the local boundary value problem (7.6).

Many algorithms may be used to solve only first-order differential equation systems. From the mathematical point of view, no harm is done by this restriction, because every higher-order system may be transformed into a first order system by the following substitutions:

$$\begin{aligned} \boldsymbol{u}_0 &\stackrel{\text{def}}{=} \boldsymbol{y} \\ \boldsymbol{u}_1 &\stackrel{\text{def}}{=} \dot{\boldsymbol{y}} \\ \ldots &\stackrel{\text{def}}{=} \ldots \\ \boldsymbol{u}_i &\stackrel{\text{def}}{=} \boldsymbol{y}^{(i)} \\ \ldots &\stackrel{\text{def}}{=} \ldots \\ \boldsymbol{u}_{n-1} &\stackrel{\text{def}}{=} \boldsymbol{y}^{(n-1)} \ . \end{aligned} \tag{7.7}$$

These transformations allow it to set up the following first order system of differential equations:

$$\begin{aligned}
\dot{\boldsymbol{u}}_0 &= \boldsymbol{u}_1 \\
\ldots &= \ldots \\
\dot{\boldsymbol{u}}_i &= \boldsymbol{u}_{i+1} \\
\ldots &= \ldots \\
\dot{\boldsymbol{u}}_{n-2} &= \boldsymbol{u}_{n-1} \\
\dot{\boldsymbol{u}}_{n-1} &= \boldsymbol{f}(t, \boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-1}) \ .
\end{aligned}$$

Obviously the system (7.1) of order n and dimension d was transformed into a system of order 1 and dimension $n\,d$. The definitions

$$\boldsymbol{u} \overset{\text{def}}{=} \begin{pmatrix} \boldsymbol{u}_0 \\ \boldsymbol{u}_1 \\ \ldots \\ \ldots \\ \boldsymbol{u}_{n-1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{F}(t, \boldsymbol{u}) \overset{\text{def}}{=} \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \ldots \\ \ldots \\ \boldsymbol{u}_{n-1} \\ \boldsymbol{f}(t, \boldsymbol{u}) \end{pmatrix} \tag{7.8}$$

allow it to write the above differential equation system in the concise matrix form:

$$\dot{\boldsymbol{u}} = \boldsymbol{F}(t, \boldsymbol{u}) \ . \tag{7.9}$$

If a dynamical system is described in this way, the designation of its solution vector $\boldsymbol{u}(t)$ as *state vector* is commonly used. If the system is specified by equations of type (7.1) of higher than the first order, its state vector must be viewed, as already mentioned, as the solution function $\boldsymbol{y}(t)$ and its first $n-1$ time derivatives.

From the mathematical point of view the systems (7.1) and (7.9) are equivalent. This is not necessarily true for their numerical solutions. From the point of view of the practitioner one should always try to solve the higher-order system, because the storage requirements and the so-called overhead of an algorithm (comprising all operations not related to the solution of the specific problem, i.e., not related to the evaluation of the right-hand sides $\boldsymbol{f}(\ldots)$ of eqns. (7.1) (or $\boldsymbol{F}(t, \boldsymbol{u})$ of eqns. (7.9))) are usually considerably smaller when sticking to the original formulation (7.1).

7.3 Euler's Algorithm

The numerical algorithms in use today to solve the initial value problem (7.5) are based on the principles outlined by Leonhard Euler in 1768 [37]. This fact is recognized by associating Euler's name with the simplest, and perhaps the most robust, integration algorithm. Euler's original analysis, in the scientific language of those days, is reproduced in Figure 7.1. In this algorithm Euler approximates the solution of the initial value problem

DE INTEGRATIONE AEQUATIONUM DIFFERENTIALIUM PER APPROXIMATIONEM

PROBLEMA 85

650. *Proposita aequatione differentiali quacunque eius integrale completum vero proxime assignare.*

SOLUTIO

Sint x et y binae variabiles, inter quas aequatio differentialis proponitur, atque haec aequatio huiusmodi habebit formam, ut sit $\frac{dy}{dx} = V$ existente V functione quacunque ipsarum x et y. Iam cum integrale completum desideretur, hoc ita est interpretandum, ut, dum ipsi x certus quidem valor, puta $x = a$, tribuitur, altera variabilis y datum quemdam valorem, puta $y = b$, adipiscatur. Quaestionem ergo primo ita tractemus, ut investigemus valorem ipsius y, quando ipsi x valor paulisper ab a discrepans tribuitur, seu posito $x = a + \omega$ ut quaeramus y. Cum autem ω sit particula minima, etiam valor ipsius y minime a b discrepabit; unde, dum x ab a usque ad $a + \omega$ tantum mutatur, quantitatem V interea tanquam constantem spectare licet. Quare posito $x = a$ et $y = b$ fiat $V = A$ et pro hac exigua mutatione habebimus $\frac{dy}{dx} = A$ ideoque integrando $y = b + A\,(x - a)$, eiusmodi scilicet constante adiecta, ut posito $x = a$ fiat $y = b$. Statuamus ergo $x = a + \omega$ fietque $y = b + A\,\omega$.

Quemadmodum ergo hic ex valoribus initio datis $x = a$ et $y = b$ proxime sequentes $x = a + \omega$ et $y = b + A\,\omega$ invenimus, ita ab his simili modo per intervalla minima ulterius progredi licet, quoad tandem ad valores a primitivis quantumvis remotos perveniatur. Quae operationes quo clarius ob oculos ponantur, sequenti modo successive instituantur.

Ipsius	valores successivi							
x	$a,$	$a',$	$a'',$	$a''',$	$a^{\mathrm{IV}},$	$\ldots,$	$'x,$	x
y	$b,$	$b',$	$b'',$	$b''',$	$b^{\mathrm{IV}},$	$\ldots,$	$'y,$	y
V	$A,$	$A',$	$A'',$	$A''',$	$A^{\mathrm{IV}},$	$\ldots,$	$'V,$	V

Scilicet ex primis $x = a$ et $y = b$ datis habetur $V = A$, tum vero pro secundis erit $b' = b + A\,(a' - a)$ differentia $a' - a$ minima pro lubitu assumpta. Hinc ponendo $x = a'$ et $y = b'$ colligitur $V = A'$ indeque pro tertiis obtinebitur $b'' = b' + A'\,(a'' - a')$, ubi posito $x = a''$ et $y = b''$ invenitur $V = A''$. ... Series autem prima valores ipsius x successivos exhibens pro lubitu accipi potest, dummodo per intervalla minima ascendat vel etiam descendat.

Fig. 7.1. Euler's method of numerical integration

$$\begin{aligned} \frac{dy}{dx} &= V(x,y) \\ y(a) &= b \end{aligned} \tag{7.10}$$

in the interval $I = [a, x]$, where x may be to the left or to the right of the initial value a of the independent argument. He subdivides the original interval I by the points a, a', a'', ..., x, where $a' - a$, $a'' - a'$, ..., are assumed to be infinitesimal, into the subintervals $I_1 = [a, a']$, $I_2 = [a', a'']$, etc. In practice these subinterval lengths will be defined as small values. The true solution is approximated by a linear function of time t within each of the subintervals $I_1 = [a, a']$, $I_2 = [a', a'']$, etc., where the linear function is defined by its value and slope at the left interval boundary (function values y and $V(x, y)$). With this procedure an approximating function is defined for all $x' \in [a, x]$, not only for the interval boundaries a, a', a'', etc. Euler's numerical solution is continuous in $I = [a, x]$, its first derivative w.r.t. the independent argument is only piece-wise continuous. Discontinuities occur at the subinterval boundaries a, a', a'', etc.

The essential elements of Euler's method are:

1. The entire integration interval $I = [a, x]$ is divided by the points $a, a', a'', \ldots,$ into subintervals I_k , $k = 1, 2, \ldots$.
2. The numerical solution is defined as a linear function of the independent argument within each of the subintervals I_k.
3. A subsidiary initial value problem is defined at the left boundary of each of the subintervals I_2 , I_3 , etc.
4. The initial values $y(a')$, $y(a'')$, etc. are defined as the numerical solution referring to the preceding subinterval I_1 , I_2 , etc., at the right interval boundaries a', a'', etc. of this subinterval.
5. The linear approximating function within each of the subintervals is defined by the initial value and the slope (right-hand side of the differential equation) at the left interval boundary.

Euler proposed his algorithm in Figure 7.1 for one scalar differential equation of order $n = 1$.

Let us transcribe Euler's algorithm to the initial value problem (7.5) in an interval $I = [t_0, t_N]$. Within this interval the initial value problem (7.5)

$$\begin{aligned} \boldsymbol{y}^{(n)} &= \boldsymbol{f}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) \\ \boldsymbol{y}^{(i)}(t_0) &\overset{\text{def}}{=} \boldsymbol{y}_0^{(i)} \ , \quad i = 0, 1, \ldots, n-1 \end{aligned}$$

is approximately solved by the following algorithm:

$$\boldsymbol{y}(t) \overset{\text{def}}{=} \boldsymbol{y}_k(t) \overset{\text{def}}{=} \sum_{i=0}^{n-1} \frac{1}{i!} (t - t_k)^i \, \boldsymbol{y}_{k0}^{(i)} + \frac{1}{n!} (t - t_k)^n \, \boldsymbol{f}\big(t_k, \boldsymbol{y}_{k0}, \dot{\boldsymbol{y}}_{k0}, \ldots, \boldsymbol{y}_{k0}^{(n-1)}\big)$$
$$t \in I_k \ , \quad k = 0, 1, \ldots, N-1 \ , \tag{7.11}$$

where the original interval $I = [t_0, t_N]$ is divided into N subintervals by the epochs (t_0), t_1, …, t_{N-1}, (t_N) (see Figure 7.2), and where the initial values within each subinterval are defined by

$$\boldsymbol{y}_{k0}^{(i)} = \begin{cases} \boldsymbol{y}_0^{(i)} & , \quad k = 0 \\ \boldsymbol{y}_{k-1}^{(i)}(t_k) \ , & \quad k > 0 \ . \end{cases} \tag{7.12}$$

The initial values are identical with those of the original problem in subinterval I_0, whereas the initial values in subinterval I_k , $k > 0$, are defined by the numerical solution $\boldsymbol{y}_{k-1}(t)$ at time t_k of the preceding interval I_{k-1}.

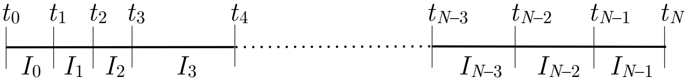


Fig. 7.2. Subdivision of the integration interval in Euler's algorithm

The algorithm (7.11, 7.12) is in the following respects a generalization of Euler's original algorithm:

- The algorithm (7.11, 7.12) is capable of dealing not only with one scalar equation, but with systems of equations. This part of the generalization was simple: Scalar coefficients had to be replaced by arrays of coefficients. Note, however, that each component of the solution may be dealt with separately (except when evaluating the right-hand sides of the differential equation systems).
- The algorithm (7.11, 7.12) is able to solve equations of higher than the first order. The numerical solution is a polynomial of degree n (Taylor series truncated after the terms of order n for differential equation systems of order n). Its $(n-1)$-st derivative (which is, e.g., needed to define the new initial values at the right interval boundaries) is, however, a linear approximation of the true $(n-1)$-st derivative of the solution.

The second generalization might have been avoided if the n-th order system would have been transformed into a first order system prior to its numerical solution. The accuracy of both approaches are comparable. The only advantage of algorithm (7.11, 7.12) over Euler's original one in Figure 7.1 resides in a reduction of the overhead of the algorithm.

The solution of the initial value problem associated with one of the intervals I_k is usually called an integration step, and the length of the subinterval I_k , $h_k \overset{\text{def}}{=} t_{k+1} - t_k$, is called the *stepsize* of the solution. The stepsize may change from step to step.

Euler's scheme clearly was meant for numerical approximations, but it proved to be very fruitful in pure mathematics, as well: Using a series of finer and finer subdivisions of the integration interval (starting from an original, arbitrary subdivision), it can be shown that the corresponding series of approximate solutions converges to the true solution. The proofs of the existence and uniqueness of solutions are, as a matter of fact, based on Euler's (numerical) solution scheme – which is truly remarkable! The convergence of Euler's solution method is very slow, however. The error decreases only linearly with the number of subintervals.

This slow convergence is the main disadvantage of Euler's scheme. It is documented by Figures 7.3 showing the error of the semi-major axis a of a hypothetical minor planet orbiting the Sun in an orbit of small eccentricity and small inclination with orbital elements defined by Table 7.1. The equations of motion of the two-body problem (4.1) were directly integrated with program NUMINT, once with the Euler method as explained above, once with the classical Runge-Kutta method as it will be outlined in section 7.4.4. The constant stepsize of the integration was one day in both cases, the integration covered a time interval of 10 years.

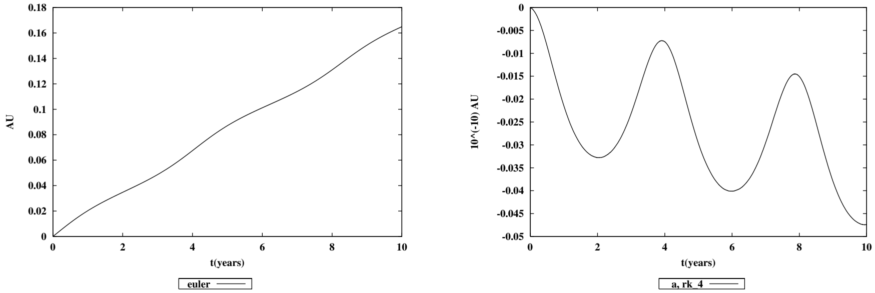


Fig. 7.3. Error of semi-major axis a when integrating an orbit of small eccentricity in a two-body potential over 10 years with a stepsize of one day; Euler method (left), Runge-Kutta of order 4 (right) (note scale differences)

After the numerical solution of the two-body problem, the orbital elements were computed for each day from the numerically integrated position and velocity vectors. As the orbital elements should be constant in the case of the two-body problem, Figure 7.3 shows directly the errors introduced by the integration. The results obviously are far from optimal in both cases. There is, however, a striking accuracy difference in the two figures: Whereas the error in a is of the order of a few hundredths of an astronomical unit (AU) when using the Euler method, the error is only of the order of a few 10^{-12} AU when using the Runge-Kutta method. The accuracy gain when using the Runge-Kutta method instead of the Euler method thus is about a factor of 10^{10}. One must admit, on the other hand, that the comparison in Figure 7.3 is not really fair: The number of evaluations of the right-hand sides of the differential equation systems is four times higher in the case of the Runge-Kutta method than in the case of the Euler method. So, with the same computational effort it would be possible to reduce the step size for the Euler method by a factor of four. This would, however, only improve the accuracy by a factor of four.

Table 7.1. Osculating elements of a "minor planet" at t_0=January 1, 2000, 0^{h}

Element	Value	Element	Value
a	2.502 AU	e	0.05
i	10°	Ω	130°
ω	30°	T_0	t_0

In summary, we may state that the Euler method is not a good choice, because of efficiency considerations. Apart from that it meets all the requirements specified in the introduction: Euler's algorithm provides an approximating function in the entire integration interval, and its error at an arbitrary epoch t is known to decrease linearly with the number of subintervals, a fact that might be exploited for an error assessment (see section 7.7 for a treatment of this topic).

7.4 Solution Methods in Overview

Before discussing in detail the collocation and multistep methods, which are fundamental in Celestial Mechanics, it is worthwhile to outline the principles underlying the major integration methods in use today. Only the principles and a few numerical examples are provided in this section. For detailed explanations of the collocation and multistep methods the readers are referred to the following sections, for extensive numerical comparisons of methods in the field of Celestial Mechanics to [75].

7.4.1 Collocation Methods

The collocation method solving the initial value problem (7.5) is in all but one aspects identical with the Euler algorithm (7.11, 7.12): Collocation algorithms approximate the initial value problem within the subintervals I_k (see Figure 7.2) by a polynomial of degree q, which is (in general) higher than in Euler's algorithm. (For $q = n$ the collocation method is reduced to the Euler algorithm.) The polynomial degree $q \geq n$ is also called the *order* of the method. The interval subdivision and the definition of the subsidiary initial value problems at the left interval boundaries are done exactly as in the Euler algorithm (using, however, usually the collocation method of order $q > n$ of the previous subinterval to define the new initial values).

The initial value problem referring to the interval I_k , $k \in \{0, 1, \ldots, N-1\}$ may be written as:

$$\begin{aligned} \boldsymbol{y}_k^{(n)} \quad &= \boldsymbol{f}\big(t, \boldsymbol{y}_k, \dot{\boldsymbol{y}}_k, \ddot{\boldsymbol{y}}_k, \ldots, \boldsymbol{y}_k^{(n-1)}\big) \\ \boldsymbol{y}_k^{(i)}(t_k) &\overset{\text{def}}{=} \boldsymbol{y}_{k0}^{(i)} \ , \quad i = 0, 1, \ldots, n-1 \ , \end{aligned} \tag{7.13}$$

where the initial values $\boldsymbol{y}_{k0}^{(i)}$ are defined by eqns. (7.12), i.e., exactly like in Euler's method.

The collocation algorithm of order $q \geq n$ approximates the initial value problem (7.13) or the boundary value problem (7.6) in the interval $I_k = [t_k, t_{k+1}]$ by a polynomial of degree q

$$\boldsymbol{y}_k(t) \overset{\text{def}}{=} \sum_{l=0}^{q} \frac{1}{l!} (t - t_k)^l \, \boldsymbol{y}_{k0}^{(l)} \ , \tag{7.14}$$

where the coefficients $\boldsymbol{y}_{k0}^{(l)}$, $l = 0, 1, \ldots, q$, are obtained by the requesting that

(a) the numerical solution assumes the initial values (7.12) and that

(b) the numerical solution solves the differential equation system at exactly $q+1-n$ different epochs t_{k_j} , $j = 1, 2, \ldots, q+1-n$, within the interval I_k (see Figure 7.4).

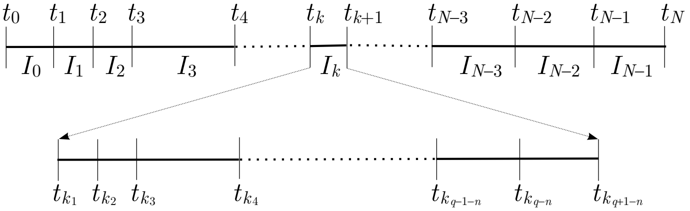


Fig. 7.4. Subdivision of the integration interval I_k for collocation algorithm

The conditions (a) are "automatically" met (replace $\boldsymbol{y}_k^{(i)}(t_k)$ in the equations for the initial values of the problem definition (7.13) using the right-hand sides of the defining eqns. (7.14) of the numerical solution). Conditions (b) are obtained explicitly by replacing $\boldsymbol{y}_k(t)$ (and its derivatives) in the differential equation system (7.13) by eqns. (7.14) for the epochs t_{k_j}:

$$\sum_{l=n}^{q} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!} \boldsymbol{y}_{k0}^{(l)} = \boldsymbol{f}\big(t_{k_j}, \boldsymbol{y}_k(t_{k_j}), \dot{\boldsymbol{y}}_k(t_{k_j}), \ldots, \boldsymbol{y}_k^{(n-1)}(t_{k_j})\big) \qquad j = 1, 2, \ldots, q+1-n \ . \tag{7.15}$$

The condition equations (7.15) are algebraic and in general non-linear in the unknowns $\boldsymbol{y}_{k0}^{(l)}$, $l = n, n+1, \ldots, q$, (because the unknowns implicitly also show up on the right-hand sides of eqns. (7.15) – the terms $\boldsymbol{y}_k^{(i)}(t_{k_j})$ must be replaced by the right-hand sides of eqns. (7.14)). Observe that the number of unknowns equals the number of condition equations.

The task of developing the explicit collocation method is rather straight forward from now on: We simply have to explain how to solve the system (7.15) of non-linear algebraic equations! The collocation method of order q is thus reduced to the problem of determining the coefficients $\boldsymbol{y}_{k0}^{(l)}$, $l = n, n{+}1, \ldots, q$, of the numerical solution (7.14). This program will be performed in section 7.5, where we will show that the above condition equations may be solved using standard methods of linear algebra.

The order q of the collocation method is defined by the program user. Orders up to about $q = 10$ to $q = 14$ make sense in a double precision floating point environment (see section 7.7).

7.4.2 Multistep Methods

When introducing the Euler method in section 7.3 and the collocation method in the section 7.4.1 it was assumed that the subintervals I_k, $k = 0, 1, \ldots, N-1$, do not overlap. There is no obvious reason for this restriction. The subintervals for collocation methods are now defined in the following, more general way:

$$I_k \overset{\text{def}}{=} [t_{k-\tilde{q}_1}, t_{k+\tilde{q}_2}] \ , \quad \tilde{q}_1, \tilde{q}_2 \geq 0 \ , \tag{7.16}$$

where $\tilde{q}_i$, $i = 1, 2$, are positive integers (natural numbers). Obviously, the case $\tilde{q}_1 = \tilde{q}_2 = 0$ does not make sense. The definition (7.16) guarantees that t_k is contained in the interval I_k.

The definition (7.16) formally does not affect the original subdivision of the integration interval by the epochs t_k. It only affects the subinterval in which the collocation epochs t_{k_j} have to lie. Many "different", potentially very powerful, collocation methods may be derived using the generalized scheme (7.16).

Definition (7.16) obviously contains the “pure” collocation method, as introduced above, as that special case for which $\tilde{q}_1 \overset{\text{def}}{=} 0$ and $\tilde{q}_2 \overset{\text{def}}{=} 1$ (and thus $I_k \overset{\text{def}}{=} [t_k, t_{k+1}]$).

For *multistep methods* the subinterval I_k is defined *either* by

$$I_k \overset{\text{def}}{=} [t_{k-q+n}, t_k] \ , \tag{7.17}$$

i.e., by $\tilde{q}_1 \overset{\text{def}}{=} q - n$ and $\tilde{q}_2 \overset{\text{def}}{=} 0$ *or* by

$$I_k \overset{\text{def}}{=} [t_{k-q+n-1}, t_{k+1}] \ , \tag{7.18}$$

i.e., by $\tilde{q}_1 \overset{\text{def}}{=} q + 1 - n$ and $\tilde{q}_2 \overset{\text{def}}{=} 1$.

Multistep methods based on the subinterval definition (7.17) are called *extrapolation methods*, those based on the subinterval definition (7.18) *interpolation methods.* The name *extrapolation method* is explained by the fact that extrapolation methods have to *extrapolate* the initial values to the epoch t_{k+1} for solving the initial value problem of the next subinterval I_{k+1}. Interpolation methods require a previous extrapolation step. They iteratively improve the solution values at t_{k+1} obtained by the extrapolation method solving the initial value problem at t_k.

The principle of multistep methods, in particular the transition from subinterval I_k to subinterval I_{k+1}, is illustrated by Figure 7.5: The interval subdivision of the original interval I is illustrated by the top figure. The second figure from the top shows the subinterval I_k for the extrapolation step. The third figure from the top shows the same subinterval I_k, indicating that the collocation epochs t_{k_j} , $j = 1, 2, \ldots, q + 1 - n$ of the extrapolation step are identical with the partition epochs t_{k-j+1} , $j = 1, 2, \ldots, q + 1 - n$ of the interval I (see top Figure). The extrapolation method in the interval I_k solves the initial value problem at epoch t_k; the differential equation system is “exactly” solved at the collocation epochs $t_{k_j} \overset{\text{def}}{=} t_{k+1-j}$, $j = 1, 2, \ldots, q + 1 - n$. The same figure also indicates that the extrapolation step has to be concluded by extrapolating the solution vector (based on the polynomial of degree q) to t_{k+1}. This extrapolation is indicated by the cross “×” in this figure. The next figure may be understood as the subinterval $I_{k,\text{inter}}$ of the interpolation step(s) related to the initial value problem referring to t_k *or* (already) as the next subinterval I_{k+1} associated with the initial value problem at t_{k+1}. If there are no interpolation steps (pure extrapolation method), the extrapolated function values at t_{k+1} are taken as the new initial conditions at t_{k+1}. If interpolation steps follow the extrpolation step, the solution (and its $n - 1$ first derivatives) of the interpolation step at t_{k+1} serve as initial values at t_{k+1}. The bottom figure shows the subinterval I_{k+1} for the extrapolation method.

Apart from using overlapping subintervals I_k, the multistep procedures have another interesting and attractive property, which distinguishes them from a

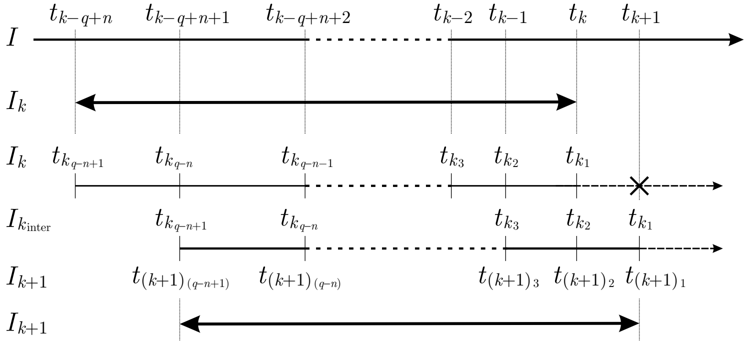


Fig. 7.5. Subdivision of the integration intervals I_k, I_{k+1} for multistep algorithm

general collocation method: With the exception of the initial value problem referring to t_0, the right-hand sides of the condition equations (7.15) are taken over as known from the previous integration steps *without ever being recomputed*! This implies that the extrapolation methods compute the right-hand sides of the differential equation systems exactly once per integration step referring to I_k, namely at t_{k+1} (indicated by the cross × in Figure 7.5). This makes multistep methods (potentially) very efficient. If one interpolation step is performed, the number of function evaluations is doubled, because the function $\boldsymbol{f}(t_{k+1})$ has to be evaluated one more time, but this still promises a high computational efficiency.

If multistep procedures are understood, as they should, as special cases of collocation methods, the designation "multistep" is somewhat misleading and not really justified. The name can only be understood, if the integration step is seen as the transition from the old initial epoch t_k to the new one at t_{k+1}. It would have been much better to call such methods collocation algorithms with overlapping intervals. When understanding the multistep methods as special cases of collocation methods it is, e.g., not necessary to lose a word about constructing a special initial sequence of function values – perhaps even, *horribile dictu*, with Runge-Kutta procedures of low order, as recommended by some experts in the field.

High-order multistep procedures ($10 \leq q \leq 14$) are extremely efficient and accurate, probably the best of all methods available today, if the stepsize $h_k \overset{\text{def}}{=} t_{k+1} - t_k$ needs not be changed frequently. Multistep methods will be discussed in detail in section 7.5.6.

In literature multistep methods are often introduced in a different way, where it is (implicitly) assumed that a prediction is possible by making not only use of the condition equations (7.15), but *in addition* of the values $\boldsymbol{y}_k(t_{k-j})$ of the solution vector $\boldsymbol{y}_k(t)$ at the previous epochs t_j , $j = 1, 2, \ldots, q-n$. Such methods in general cannot be stable. In the author's opinion they do not deserve the name "integration method", which is why they are not considered here. In the best case, they have a role to play when studying the asymptotic behavior of differential equation systems. More information may, e.g., be found in [53].

7.4.3 Taylor Series Methods

Exactly as the collocation methods, the Taylor series algorithms differ only in one aspect from Euler's scheme (7.11, 7.12): The Taylor series development is truncated only after the terms of order $q > n$ and not already after the terms of order n. The derivatives of the orders n to q have to be computed with the help of the differential equation system:

$$\begin{aligned}\boldsymbol{y}_k(t) &= \sum_{l=0}^{q} \frac{1}{l!}\,(t-t_k)^l\,\boldsymbol{y}_{k0}^{(l)} \\ &= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t-t_k)^l\,\boldsymbol{y}_{k0}^{(l)} + \sum_{l=n}^{q} \frac{1}{l!}\,(t-t_k)^l\,\boldsymbol{f}_{k0}^{(l-n)}\ ,\end{aligned} \tag{7.19}$$

where the derivatives on the right-hand side all refer to the epoch t_k, and where

$$\boldsymbol{f}_{k0} \overset{\text{def}}{=} \boldsymbol{f}\big(t_k, \boldsymbol{y}_{k0}, \dot{\boldsymbol{y}}_{k0}, \ldots, \boldsymbol{y}_{k0}^{(n-1)}\big)\ . \tag{7.20}$$

The first derivative of $\boldsymbol{f}_{k0} \overset{\text{def}}{=} \boldsymbol{f}(t_k)$ is computed as follows:

$$\begin{aligned}\boldsymbol{y}_k^{(n+1)}(t_k) \overset{\text{def}}{=} \boldsymbol{y}_{k0}^{(n+1)} &= \dot{\boldsymbol{f}}_{k0} \\ &= \frac{\partial \boldsymbol{f}_{k0}}{\partial t} + \sum_{i=0}^{n-1}\sum_{l=1}^{d} \frac{\partial \boldsymbol{f}_{k0}}{\partial y_{k0_l}^{(i)}}\, y_{k0_l}^{(i+1)} \\ &= \frac{\partial \boldsymbol{f}_{k0}}{\partial t} + \sum_{i=0}^{n-2}\sum_{l=1}^{d} \frac{\partial \boldsymbol{f}_{k0}}{\partial y_{k0_l}^{(i)}}\, y_{k0_l}^{(i+1)} + \sum_{l=1}^{d} \frac{\partial \boldsymbol{f}_{k0}}{\partial y_{k0_l}^{(n-1)}}\, f_{k0_l}\ ,\end{aligned} \tag{7.21}$$

where the index l characterizes the component no l of the corresponding array. Higher-order derivatives are obtained by taking the derivatives of eqns. (7.21).

There can be no doubt that an algorithm resulting from such a scheme would be much better than the Euler algorithm (7.11, 7.12) and that it would meet all the requirements postulated in the introductory section 7.1. The result of the integration is in particular an approximating function which may be used everywhere within the interval considered. Also, local error control could be easily established.

A practical consideration destroys the concept: The concrete form of the derivatives of $\boldsymbol{f}(t)$ is problem-dependent and the resulting expressions may become close to unmanageable. This is true for most problems in Celestial

Mechanics. Efficient Taylor series methods may, however, be developed for the solution of linear differential equation systems. The issue will be taken up again in section 7.6.

Taylor Series Solution for the Two-Body Problem. In order to demonstrate the complexity of Taylor series methods we include this type of algorithm to solve the two-body problem. With the advent of digital computers quite a few of these algorithms were developed. We quote the algorithm given by A. E. Roy in [94]. The algorithm solves the following initial value problem:

$$\begin{aligned} \ddot{\boldsymbol{r}} &= -\frac{\boldsymbol{r}}{r^3} \\ \boldsymbol{r}(t_0) &\overset{\text{def}}{=} \boldsymbol{r}_0 \\ \dot{\boldsymbol{r}}(t_0) &\overset{\text{def}}{=} \dot{\boldsymbol{r}}_0 \ . \end{aligned} \tag{7.22}$$

Note that the gravity constant was put to $\mu = 1$ which has implications on the units used. The approximate solution is sought in the form of a power polynomial in $t - t_0$ (which is equivalent to a Taylor series):

$$\boldsymbol{r}(t) = \sum_{i=0}^{q} \boldsymbol{r}_i \, (t - t_0)^i \ . \tag{7.23}$$

With the introduction of the four scalar auxiliary variables (for better reference the notation is taken over from [94] and only used in the context of this particular algorithm)

$$\begin{aligned} u(t) &\overset{\text{def}}{=} \frac{1}{r^3} = \sum_{i=0}^{q} u_i \, (t - t_0)^i \\ w(t) &\overset{\text{def}}{=} \frac{1}{r^2} = \sum_{i=0}^{q} w_i \, (t - t_0)^i \\ s(t) &\overset{\text{def}}{=} \boldsymbol{r} \cdot \dot{\boldsymbol{r}} = \sum_{i=0}^{q} s_i \, (t - t_0)^i \\ \sigma(t) &\overset{\text{def}}{=} s\, w = \sum_{i=0}^{q} \sigma_i \, (t - t_0)^i \ , \end{aligned} \tag{7.24}$$

where the functions u and w obey the differential equations

$$\begin{aligned} \dot{u} &= -\,3\,u\,\sigma \\ \dot{w} &= -\,2\,u\,\sigma \end{aligned} \tag{7.25}$$

and the original differential equation may be written as

$$\ddot{\boldsymbol{r}} = -\,u\,\boldsymbol{r}, \tag{7.26}$$

it is possible to calculate the terms of all power polynomials recursively:

$$
\begin{aligned}
u_0 &= \frac{1}{r_0^3} \\
w_0 &= \frac{1}{r_0^2} \\
s_0 &= \boldsymbol{r}_0 \cdot \dot{\boldsymbol{r}}_0 \\
\sigma_0 &= w_0 \, s_0 \\
\boldsymbol{r}_{j+2} &= -\frac{1}{(j+1)(j+2)} \sum_{i=0}^{j} u_{j-i} \, \boldsymbol{r}_i \\
u_{j+1} &= -\frac{3}{j+1} \sum_{i=0}^{j} u_i \, \sigma_{j-i} \\
w_{j+1} &= -\frac{2}{j+1} \sum_{i=0}^{j} w_i \, \sigma_{j-i} \\
s_{j+1} &= \sum_{i=0}^{j+1} (i+1) \, \boldsymbol{r}_{i+1} \cdot \boldsymbol{r}_{j-i+1} \\
\sigma_{j+1} &= \sum_{i=0}^{j+1} w_i \, s_{j-i+1} \\
& \quad j = 0, 1, \ldots, q-2 \; .
\end{aligned}
\tag{7.27}
$$

Note that the algorithm (7.27) is valid for all conic sections. When generating a computer code of the above algorithm, the vector $\boldsymbol{r}(t)$ may be treated as two dimensional in the orbital plane (with the first axis pointing from the focal point of the conic section to the pericenter). The algorithm only needs the eccentricity e, the true anomaly v and (of course) the maximum order of the development as input. The algorithm can be modified not to return the polynomial coefficients $\boldsymbol{r}_i$, but the values of the derivatives at t_0, $\boldsymbol{r}^{(i)}(t_0)$. For a "real" orbit, characterized, e.g., in addition by a, the term $\boldsymbol{r}^{(i)}$ has to be scaled (multiplied) by an^i (for elliptical orbits), where n is the mean motion.

7.4.4 Runge-Kutta Methods

It is logical to deal with Runge-Kutta methods immediately after the Taylor series method. Runge-Kutta methods of the order q compute the value of the solution function (and its first $n-1$ derivatives) for exactly one instant near the initial epoch, namely at $t = t_k + h_k$, with an accuracy equivalent to a Taylor series up to order q. Runge-Kutta methods are named after two German mathematicians, Carl David Tolmé Runge (1856–1927) and Martin

Wilhelm Kutta (1867–1944), who developed a procedure of order $q = 4$ for differential equation systems of order $n = 1$. Runge-Kutta algorithms are robust and simple in application. They are very popular and therefore covered in virtually every textbook of numerical analysis and in many textbooks of Celestial Mechanics. As the result of Runge-Kutta procedures is not an approximating function as requested in section 7.1, we only develop the key ideas and reproduce the best-known algorithm. Runge-Kutta procedures of the orders $q = 4, 7$, and 8 are available in the program NUMINT (see Chapter II- 6 of Part III).

Let us approximate the solution of the initial value problem

$$\begin{aligned} \dot{\boldsymbol{y}} \quad &= \boldsymbol{f}(t, \boldsymbol{y}) \\ \boldsymbol{y}(t_0) &\overset{\text{def}}{=} \boldsymbol{y}_0 \end{aligned} \tag{7.28}$$

of a differential equation system of first order at $t = t_0 + h$ by an expression which is equivalent to a Taylor series up to degree and order q

$$\begin{aligned} \boldsymbol{y}(t_0 + h) &= \sum_{l=0}^{q} \frac{1}{l!} \boldsymbol{y}_0^{(l)} h^l + \boldsymbol{O}\left(h^{q+1}\right) \\ &= \boldsymbol{y}_0 + \sum_{l=1}^{q} \frac{1}{l!} \boldsymbol{f}^{(l-1)}(t_0) h^l + \boldsymbol{O}\left(h^{q+1}\right) . \end{aligned} \tag{7.29}$$

Runge-Kutta algorithms do not determine (approximations for) the coefficients of the Taylor series. They merely provide the sum on the right-hand side of eqn. (7.29) to the accuracy specified.

The first derivative of $\boldsymbol{f}(t)$ at $t = t_0$, namely $\dot{\boldsymbol{f}}(t_0) = \dot{\boldsymbol{f}}_0$, may be written as

$$\begin{aligned} \dot{\boldsymbol{f}}_0 &= \left\{ \frac{\partial}{\partial t} + \dot{\boldsymbol{y}}_0 \cdot \frac{\partial}{\partial \boldsymbol{y}_0} \right\} \boldsymbol{f}(t_0) \\ &= \left\{ \frac{\partial}{\partial t} + \boldsymbol{f}_0 \cdot \frac{\partial}{\partial \boldsymbol{y}_0} \right\} \boldsymbol{f}_0 \\ &= \mathrm{D}\boldsymbol{f}_0 , \end{aligned} \tag{7.30}$$

where the scalar operator D is explicitly defined as

$$\mathrm{D} \overset{\text{def}}{=} \left\{ \frac{\partial}{\partial t} + \boldsymbol{f} \cdot \left(\frac{\partial}{\partial \boldsymbol{y}} \right) \right\}_{t_0} = \left\{ \frac{\partial}{\partial t} + \sum_{j=1}^{d} f_j \left(\frac{\partial}{\partial y_j} \right) \right\}_{t_0} . \tag{7.31}$$

This definition allows it to write the higher-order derivatives in a very concise way:

$$\begin{aligned} \dot{\boldsymbol{f}}_0 &= \mathrm{D}\boldsymbol{f}_0 \\ \ddot{\boldsymbol{f}}_0 &= \mathrm{D}^2\boldsymbol{f}_0 \\ \ldots &= \ldots \\ \boldsymbol{f}_0^{(q)} &= \mathrm{D}^q\boldsymbol{f}_0 \; . \end{aligned} \tag{7.32}$$

The actual computation of the operator D^i becomes more and more cumbersome with increasing order i, mainly due to the fact that $\boldsymbol{f}_0$ must be considered as a function of t and $\boldsymbol{y}_0$ in the operator D defined by (7.31). If we replace the functions $\boldsymbol{f}_0$ by their values $\boldsymbol{\nu}_0$ at $t = t_0$, we obtain the following linear operator:

$$\mathrm{D}_0 \overset{\text{def}}{=} \frac{\partial}{\partial t} + \boldsymbol{\nu}_0 \cdot \frac{\partial}{\partial \boldsymbol{y}} \; . \tag{7.33}$$

With this linear operator (7.33) the derivatives (7.32) at t_0 may be written as follows:

$$\begin{aligned} \dot{\boldsymbol{f}}_0 &= \mathrm{D}_0\boldsymbol{f}_0 \\ \ddot{\boldsymbol{f}}_0 &= \mathrm{D}_0^2\boldsymbol{f}_0 + (\nabla_y \mathrm{D}_0)\boldsymbol{f}_0 \\ \boldsymbol{f}_0^{(3)} &= \mathrm{D}_0^3\boldsymbol{f}_0 + \ldots \\ \ldots &= \ldots \; . \end{aligned} \tag{7.34}$$

With eqns. (7.34) and $h \overset{\text{def}}{=} t - t_0$ for the time argument relative to t_0 the Taylor series expansion (7.29) may be written as

$$\boldsymbol{y}(t) = \boldsymbol{y}_0 + h\,\boldsymbol{f}_0 + \frac{h^2}{2!}\mathrm{D}_0\boldsymbol{f}_0 + \frac{h^3}{3!}\left\{\mathrm{D}_0^2\boldsymbol{f}_0 + (\nabla_y \mathrm{D}_0)\boldsymbol{f}_0\right\} + \ldots \; . \tag{7.35}$$

It is the goal of Runge-Kutta algorithms of order q to approximate the expansion (7.35) up to terms of order q by a linear combination of a (minimum) number of function values $\boldsymbol{f}(t_i, \boldsymbol{y})$ at different locations in the vicinity of the point $(t_0, \boldsymbol{y}_0)$ in the $(d+1)$-dimensional space. The general explicit Runge-Kutta algorithm reads as:

$$\begin{aligned} \boldsymbol{k}_1 &= h\,\boldsymbol{f}(t_0, \boldsymbol{y}_0) \\ \boldsymbol{k}_2 &= h\,\boldsymbol{f}(t_0 + \alpha\, h, \boldsymbol{y}_0 + \beta\, \boldsymbol{k}_1) \\ \boldsymbol{k}_3 &= h\,\boldsymbol{f}(t_0 + \alpha_1 h, \boldsymbol{y}_0 + \beta_1 \boldsymbol{k}_1 + \gamma_1 \boldsymbol{k}_2) \\ \ldots &= \ldots \\ \boldsymbol{k}_m &= c_1\,\boldsymbol{k}_1 + c_2\,\boldsymbol{k}_2 + c_3\,\boldsymbol{k}_3 + \ldots \\ \boldsymbol{y}(t_0 + h) &= \boldsymbol{y}_0 + \boldsymbol{k}_m \; , \end{aligned} \tag{7.36}$$

where the constants $\alpha,\ \beta, \ldots,\ c_1,\ c_2, \ldots$, have to be defined in such a way that the last equation in formulae (7.36) is equivalent to the Taylor series (7.35) truncated after terms of order q. In order to obtain the condition equations for the coefficients, the right-hand sides of eqns. (7.36) have to be developed up to the required order relative to the point $(t_0, \boldsymbol{y}_0)$. This is achieved by the following formal development:

$$\boldsymbol{f}(t_0 + \Delta t, \boldsymbol{y}_0 + \Delta \boldsymbol{y}) = \sum_{l=0}^{m} \frac{1}{l!} \tilde{\mathrm{D}}^l \boldsymbol{f} \ , \tag{7.37}$$

where $\tilde{\mathrm{D}}$ is a linear operator which differs from the linear operator D_0 defined by eqn. (7.33) only by the coefficients of the partial derivatives:

$$\tilde{\mathrm{D}} \overset{\text{def}}{=} \Delta t \, \frac{\partial}{\partial t} + \Delta \boldsymbol{y} \cdot \frac{\partial}{\partial \boldsymbol{y}} \ . \tag{7.38}$$

From here onwards, the procedure is in principle clear: The right-hand sides of the eqns. (7.36) have to be developed using the expansion (7.37), and the result has to be identical with the truncated Taylor series (7.35). This comparison results in condition equations for the coefficients. The actual computations are rather elaborate, in particular for orders $q > 4$.

The best-known Runge-Kutta algorithm undoubtedly is that of order $q = 4$, probably reproduced in every treatment of numerical analysis:

$$\begin{aligned}
\boldsymbol{k}_1 \quad &= h\,\boldsymbol{f}(t_0, \boldsymbol{y}_0) \\
\boldsymbol{k}_2 \quad &= h\,\boldsymbol{f}\left(t_0 + \tfrac{1}{2}\,h, \boldsymbol{y}_0 + \tfrac{1}{2}\,\boldsymbol{k}_1\right) \\
\boldsymbol{k}_3 \quad &= h\,\boldsymbol{f}\left(t_0 + \tfrac{1}{2}\,h, \boldsymbol{y}_0 + \tfrac{1}{2}\,\boldsymbol{k}_2\right) \\
\boldsymbol{k}_4 \quad &= h\,\boldsymbol{f}\big(t_0 + h, \boldsymbol{y}_0 + \boldsymbol{k}_3\big) \\
\ldots \quad &= \ldots \\
\boldsymbol{k}_m \quad &= \tfrac{1}{6}\left(\boldsymbol{k}_1 + 2\,\boldsymbol{k}_2 + 2\,\boldsymbol{k}_3 + \boldsymbol{k}_4\right) \\
\ldots \quad &= \ldots \\
\boldsymbol{y}(t_0 + h) &= \boldsymbol{y}_0 + \boldsymbol{k}_m \ .
\end{aligned} \tag{7.39}$$

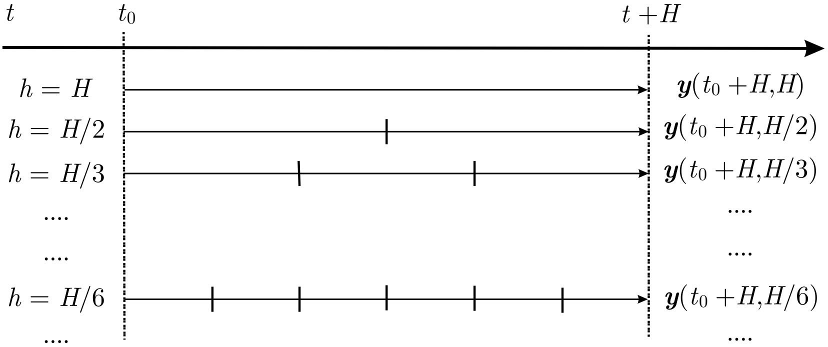
The idea underlying Runge-Kutta procedures is attractive from the mathematical point of view, and the resulting algorithms are not only powerful, but also easy to use. Due to the heavy algebra involved in the explicit computation of the coefficients, it took quite some time to develop higher-order methods, methods for higher-order equations, and algorithms providing error control. Such algorithms are available today up to about order $q = 12$ for first and second order differential equation systems. We refer to the pioneer work [38] by E. Fehlberg, and to [75] for an overview of more recent developments applied to problems of Celestial Mechanics.

7.4.5 Extrapolation Methods

According to Gear [45] extrapolation methods are based on an idea originally developed by L. F. Richardson in the early 20th century. The method is also called "deferred approach to the limit". The key idea is illustrated by Figure 7.6 and may be summarized as follows: The solution $\boldsymbol{y}(t)$ of the initial value problem

$$\begin{aligned} \dot{\boldsymbol{y}} &= \boldsymbol{f}(t, \boldsymbol{y}) \\ \boldsymbol{y}(t_0) &\overset{\text{def}}{=} \boldsymbol{y}_0 \end{aligned} \tag{7.40}$$

at one and the same epoch $t = t_0 + H$ is calculated using one and the same simple algorithm (e.g., Euler's method), however, with smaller and smaller (constant) stepsizes h_j , e.g., $h_j \overset{\text{def}}{=} \frac{H}{j}$, in subsequent approximations.



t
t_0
$t + H$
$h = H$
$h = H/2$
$h = H/3$
....
....
$h = H/6$
....
$\boldsymbol{y}(t_0 + H, H)$
$\boldsymbol{y}(t_0 + H, H/2)$
$\boldsymbol{y}(t_0 + H, H/3)$
....
....
$\boldsymbol{y}(t_0 + H, H/6)$
....

Fig. 7.6. Principles of extrapolation methods

Denoting the approximation of the solution at $t_0 + H$ with stepsize h_j by $\boldsymbol{y}(t_0+H, h_j)$, one may then interpret these solutions at $t = t_0+H$ as functions of the stepsizes h_j. If all in all $q + 1$ approximations $\boldsymbol{y}(t_0 + H, h_j)$, $j = 1, 2, \ldots, q+1$, were calculated at $t_0 + H$, it is possible to represent them by a polynomial of degree q in the stepsize h and to *use the value of the polynomial at* $h = 0$ *as the estimate for the solution at* $t_0 + H$. It is intuitively clear that this approximation corresponds to a hypothetical stepsize of $h = 0$.

Let us develop the procedure more explicitly using the Euler method as the underlying integration method. The numerical solution of the initial value problem (7.40) at $t_0 + H$ may thus be written as a function of the stepsize h of the Euler integration steps used, where the function values for $h = h_j \overset{\text{def}}{=} \frac{H}{j}$, $j = 1, 2, \ldots$, are computed as:

$$\begin{aligned}
h_1 = H:\ & \boldsymbol{y}(t_0+H,h_1) &&= \boldsymbol{y}_0 + h_1\,\boldsymbol{f}(t_0,\boldsymbol{y}_0)\\
h_2 = \tfrac{H}{2}:\ & \boldsymbol{y}(t_0+h_2,h_2) &&= \boldsymbol{y}_0 + h_2\,\boldsymbol{f}(t_0,\boldsymbol{y}_0)\\
& \boldsymbol{y}(t_0+H,h_2) &&= \boldsymbol{y}(t_0+h_2,h_2) + h_2\boldsymbol{f}\big(t_0+h_2,\boldsymbol{y}(t_0+h_2,h_2)\big)\\
h_3 = \tfrac{H}{3}:\ & \boldsymbol{y}(t_0+h_3,h_3) &&= \boldsymbol{y}_0 + h_3\,\boldsymbol{f}(t_0,\boldsymbol{y}_0)\\
& \boldsymbol{y}(t_0+2\,h_3,h_3) &&= \boldsymbol{y}(t_0+h_3,h_3) + h_3\boldsymbol{f}\big(t_0+h_3,\boldsymbol{y}(t_0+h_3,h_3)\big)\\
& \boldsymbol{y}(t_0+H,h_3) &&= \boldsymbol{y}(t_0+2h_3,h_3) + h_3\boldsymbol{f}\big(t_0+2\,h_3,\boldsymbol{y}(t_0+2\,h_3,h_3)\big)\\
\ldots & && .
\end{aligned} \tag{7.41}$$

The numerical solutions $\boldsymbol{y}(t_0+H,h)$ at epoch $t = t_0+H$ may be represented by a special polynomial of degree q (plus an error term of order $q+1$ in h) in the (variable) stepsize h:

$$\begin{aligned}
\boldsymbol{y}(t_0+H,h) &\overset{\text{def}}{=} \sum_{i=0}^{q} \boldsymbol{c}_i\,h^i \;+\; \boldsymbol{O}(h^{q+1})\\
&= \boldsymbol{y}(t_0+H,0) \;+\; \sum_{i=1}^{q} \boldsymbol{c}_i\,h^i \;+\; \boldsymbol{O}(h^{q+1})\ ,
\end{aligned} \tag{7.42}$$

where

$$\boldsymbol{c}_0 \overset{\text{def}}{=} \boldsymbol{y}(t_0+H,0) \tag{7.43}$$

has to be understood as the “true” solution vector (corresponding to stepsize $h=0$) at epoch $t = t_0+H$.

From now on, the procedure to calculate $\boldsymbol{c}_0 \overset{\text{def}}{=} \boldsymbol{y}(t_0+H,0)$ is straight forward: The coefficients $\boldsymbol{c}_i\ ,\ i=0,1,\ldots$, in formula (7.42) are determined using the approximations $\boldsymbol{y}(t_0+H,h_j)\ ,\ j=1,2,\ldots,q+1$, as function values to be assumed by the representation (7.42):

$$\sum_{i=0}^{q} h_j^i\,\boldsymbol{c}_i \overset{\text{def}}{=} \boldsymbol{y}\big(t_0+H,h_j\big)\ ,\quad j=1,2,\ldots,q+1\ . \tag{7.44}$$

As the terms $\boldsymbol{O}(h^{q+1})$ of order $q+1$ in h were neglected when setting up the condition equations (7.44), all terms $\boldsymbol{c}_i\,h_j^i$ are accurate to order q in H. By solving the linear system of equations (7.44) we obtain in particular the term $\boldsymbol{c}_0$

$$\boldsymbol{c}_0 = \boldsymbol{y}(t_0+H,0) + \boldsymbol{O}(h^{q+1})\ . \tag{7.45}$$

Extrapolation methods can be substantially improved by replacing Euler's method to generate the successive approximations at $t = t_0+H$ with the steps $h_j\ ,\ j=1,2,\ldots$, by higher-order methods with the following error property:

$$\boldsymbol{y}(t_0+H,h) - \boldsymbol{y}(t_0+H,0) = \sum_{i=1}^{n} c_i\,h^{ri} \;+\; \boldsymbol{O}(h^{n+1})\ , \tag{7.46}$$

where $r > 1$ is a positive number. The errors at $t = t_0 + H$ may thus be represented as polynomial in $\xi \overset{\text{def}}{=} h^r$, which is why the error property (7.46) is superior to the error property (7.42) associated with Euler's method.

Gragg [47] gives a simple algorithm based on a second order integrator for which $r = 2$. The individual step in formulae (7.41) for $h \in \{h_1, h_2, \ldots,\}$ is replaced by

$$\begin{aligned}
\boldsymbol{y}(t_0 + h, h) &= \boldsymbol{y}_0 \;+\; h\,\boldsymbol{f}(t_0, \boldsymbol{y}_0)\\
\boldsymbol{y}(t_0 + 2\,h, h) &= \boldsymbol{y}_0 \;+\; 2\,h\,\boldsymbol{f}\big(t_0 + h, \boldsymbol{y}(t_0 + h, h)\big)\\
\ldots &= \ldots\\
\boldsymbol{y}(t_0 + i\,h, h) &= \boldsymbol{y}\big(t_0 + (i-2)\,h, h\big) + 2h\boldsymbol{f}\big(t_0 + (i-1)\,h, \boldsymbol{y}(t_0 + (i-1)\,h, h)\big)\\
\ldots &= \ldots\\
\boldsymbol{y}(t_0 + n\,h, h) &= \boldsymbol{y}\big(t_0 + (n-2)\,h, h\big) + 2h\boldsymbol{f}\big(t_0 + (n-1)\,h, \boldsymbol{y}(t_0 + (n-1)\,h, h)\big)\\
\boldsymbol{y}(t_0 + H, h) &= \tfrac{1}{2}\Big[\boldsymbol{y}(t_0 + n\,h, h) \;+\; \boldsymbol{y}(t_0 + (n-1)\,h, h)\\
&\qquad + h\,\boldsymbol{f}\big(t_0 + n\,h, \boldsymbol{y}(t_0 + n\,h, h)\big)\Big]\ ,
\end{aligned} \tag{7.47}$$

where n must be an even number. Press et al. [88] recommend to use the sequence h_n , $n = 2, 4, 8, 10, \ldots$, with $n_{\max} \approx 16$.

The algorithm may be brought into a very efficient form based on divided difference schemes, where the approximations for $\boldsymbol{y}(t_0 + H)$ are obtained in increasing orders of h^2. The error behavior allows it easily to estimate the accuracy of the result at $t = t_0 + H$ (using in essence analogous methods as in the case of collocation algorithms). One of the most advanced algorithms was developed by J. Stoer and R. Bulirsch, see, e.g., [112] or [113]. An extrapolation method of selectable order q is available in program NUMINT.

One should be aware of the fact that extrapolation algorithms provide accurate values for the solution only at the interval boundaries. No comparable approximation is available between the interval boundaries. If the only interest in the solution of an initial value problem consists in propagating the solution vector from the initial epoch t_0 to a remote epoch t_N , extrapolation algorithms may be extremely powerful tools. If, on the other hand, an approximating function is needed, their use seems somewhat limited. Extrapolation methods certainly are superior to Runge-Kutta methods, among other because high-order extrapolation methods may be obtained easily, without the painful algebra involved in the establishment of the Runge-Kutta methods.

7.4.6 Comparison of Different Methods

Five different methods were outlined in this introductory section. They all are of orders $q > n$, higher than the order n of the differential equation system, which in turn is the order of Euler's method. There are many different ways to compare methods. In Celestial Mechanics it is a well-established practice to apply the methods to the solution of the two-body problem, where usually the distinction is made between applications to orbits of small and large eccentricities.

Figure 7.7 compares the performance of four eight-order methods, namely the Runge-Kutta, the extrapolation, the multistep, and the collocation methods. A constant stepsize of $h = 40$ days was used for the Runge-Kutta and the extrapolation method, one of $H = 40$ days for the extrapolation method, and one of $h = 40/(q-1) \approx 6$ days for the multistep method. An orbit with the small eccentricity $e = 0.05$ (initial elements defined by Table 7.1) was integrated. The integration interval covers a time period of 1000 years corresponding to 252.6 revolutions.

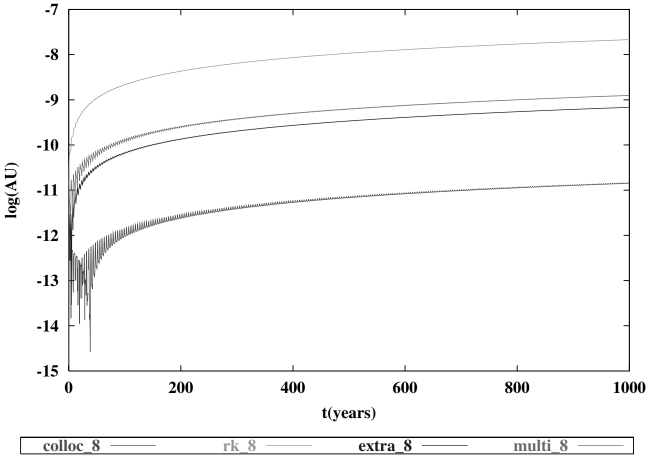


Fig. 7.7. Error of semi-major axis a when integrating a minor planet orbit with $e = 0.05$ a in two-body potential over 1000 years with a stepsize of 40 days; Collocation, Runge-Kutta, Extrapolation, and Multistep (stepsize of $h = 6$ days) (all of order 8)

The integration was performed in rectangular coordinates, i.e., the equations of motion (4.1) using $\mu = k^2$ were integrated, where these equations were transformed into a first-order system when using the Runge-Kutta and the extrapolation methods. (The integration was based on the second order equations when using the collocation and multistep methods).

At intervals of 120 days the rectangular coordinates of the position- and velocity-vectors were transformed into osculating orbital elements. The differences of these elements w.r.t. the initial elements, i.e., the integration errors in the elements, were then stored in a file. Figure 7.7 shows the logarithm of the (absolute values of) the integration errors in the semi-major axis a (the results for the other elements show in essence the same pattern).

Figure 7.7 first of all tells that all methods are excellent. The results are much better than those achieved in Figures 7.3. At the end of the integration interval (i.e., after 1000 years) the error lies between the limits 10^{-11} AU for the collocation method and 10^{-8} AU for the Runge-Kutta method of order 8. The extrapolation method and the multistep method with an accuracy of about 10^{-9} AU are about comparable.

The order of a method in essence defines the order of magnitude of the error. The efficiency of a method is defined by the number n_f of evaluations of the right-hand sides of the differential equation system per revolution (except for those cases where the evaluation of the right-hand sides of the differential equation systems is trivial). These numbers were $n_f = 470$ for the collocation method, $n_f = 759$ for the extrapolation method, $n_f = 482$ for the multistep method, and $n_f = 470$ for the Runge-Kutta method. These numbers would favor the collocation and multistep methods for this kind of problems.

One should not "over-interpret" results of the kind represented by Figure 7.7. With increasing integration order q the mutual performance may change considerably. For more information we refer to section 7.5.4.

7.5 Collocation

From the author's point of view the collocation method is the central tool for solving ordinary differential equation. Collocation provides a numerical solution of type (7.14), representing each component of the solution as a truncated Taylor series in each of the subintervals I_k.

The basic principles underlying the collocation method were already outlined in section 7.4.1. The collocation method is based on the requests that

(a) the numerical solution of the initial or boundary value problem within interval I_k assumes the same initial or boundary values as the true solution of the same problem, and that

(b) the numerical approximation solves the differential equation system at exactly $q+1-n$ different epochs within the interval I_k (see eqns. (7.15)).

The theory of collocation methods is treated in sections 7.5.1 and 7.5.2, where the solution of the initial value problem will be studied in section 7.5.1, that of the local boundary value problem in section 7.5.2. Section 7.5.3 is devoted to computational efficiency, numerical stability, and, to some extent, to the elegance of the formalism. Collocation methods are applied to the problem of integrating an orbit of small eccentricity in section 7.5.4, a very important class of problems in Celestial Mechanics. The integration interval may be divided into subintervals I_k of the same length $h \overset{\text{def}}{=} t_{k+1} - t_k$ in this case. The example demonstrates the power of the method for problems of this kind. As the analytical solution is known, we also gain a first impression of the accumulation of errors in this section. In section 7.5.5 we first show that collocation methods based on a constant stepsize h are inefficient or even inappropriate to solve the two-body problem when the orbital eccentricity e is no longer small. The remainder of the section is devoted (a) to the development of a very simple method to define the stepsize (subinterval length) $h_k \overset{\text{def}}{=} t_{k+1} - t_k$ "in real time" by using the information available from the collocation method and (b) to demonstrate the power of the automatic stepsize control. In section 7.4.2 it was already shown that multistep methods are special cases of collocation methods. In the concluding section 7.5.6 we further develop the concept.

7.5.1 Solution of the Initial Value Problem

In the case of an initial value problem, requirement (a) states that the first $n-1$ derivatives of the numerical and the true solution are identical at t_k:

$$\boldsymbol{y}_k^{(i)}(t_k) = \sum_{l=i}^{q} \frac{1}{(l-i)!}\,(t_k - t_k)^{l-i}\,\boldsymbol{y}_{k0}^{(l)} = \boldsymbol{y}_{k0}^{(i)}\;, \quad i = 0, 1, \ldots, n-1\;, \quad (7.48)$$

where the values $\boldsymbol{y}_{k0}^{(i)}$, $i = 0, 1, \ldots, n-1$, are the known initial values at time t_k (see eqn. (7.13)).

According to requirement (b) the numerical solution has to solve the differential equation system for $q+1-n$ different epochs within the interval I_k. This subdivision of the subinterval I_k is illustrated by Figure 7.4, where it was assumed that $t_{k_1} = t_k$ and $t_{k_{q+1-n}} = t_{k+1}$. We will always adopt this assumption, subsequently, would like to point out, however, that even a more general selection of the collocation epochs t_{k_j} would be allowed. The coefficients are determined by the algebraic condition equations (7.15), which are repeated here for convenience:

$$\sum_{l=n}^{q} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!}\,\boldsymbol{y}_{k0}^{(l)} = \boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k(t_{k_j}), \dot{\boldsymbol{y}}_k(t_{k_j}), \ldots, \boldsymbol{y}_k^{(n-1)}(t_{k_j})\Big)\;,$$
$$j = 1, 2, \ldots, q+1-n\;.$$

Equations (7.15) represent an algebraic system of equations for the determination of the coefficients $\boldsymbol{y}_{k0}^{(l)}$, $l = n, n+1, \ldots, q$. Note that the first n coefficients are already known from eqns. (7.48). The problem of numeric integration in the interval I_k with a high-order integrator is thus one of solving the above system of algebraic equations, where the number of equations and the number of unknowns are the same.

The left-hand sides of eqns. (7.15) are linear in the unknown coefficients. These left-hand sides may even be dealt with component by component, because in the (scalar) equation referring to a particular component, only the coefficients corresponding to this component show up. In general, the unknown coefficients are also contained on the right-hand sides of the above condition equations. Depending on the problem considered, the dependence of the function $\boldsymbol{f}(\ldots)$ on the coefficients may be quite complicated. The actual structure of the system (7.15) therefore depends on the structure of the original differential equation system (7.1).

The solution of eqns. (7.15) is reduced to the solution of d separate linear systems of equations (namely one for each component of the solution vector) if the function $\boldsymbol{f}$ does neither depend on the solution vector nor on its derivatives, i.e., if $\boldsymbol{f} \overset{\text{def}}{=} \boldsymbol{f}(t)$ is "only" an explicit function of time. This situation is encountered when solving a definite integral by numerical quadrature. Under this assumption the equations for component m read as:

$$\sum_{l=n}^{q} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!} \, y_{k0_m}^{(l)} = f_m(t_{k_j}) \ , \quad j = 1, 2, \ldots, q+1-n \ . \tag{7.49}$$

Equations (7.49) characterize virtually every (useful) formula for numerical quadrature. A unique solution exists, as long as $t_{k_i} \neq t_{k_l}$ for $i \neq l$. The condition equations of type (7.49) will be further studied in section 7.6.

If the differential equation system is linear, i.e., if it is of the form (7.4), the system of condition equations (7.15) is linear, as well. The linear algebraic system is, however, of dimension $d' = d \ (q-n)$. Depending on the specific problem, d' may assume rather high values, what might be prohibitive for the numerical solution of a linear system of equations. Many scalar linear differential equations of mathematical physics are, however, solved in a very elegant and efficient way by solving the linear system of algebraic equations. It is worthwhile to point out already at this point that *the solution of a linear differential equation system may be reduced to the solution of a linear system of algebraic equations.*

Having sorted out the two simple special cases (solution of integrals and of linear equations), we are left with the general problem, where $\boldsymbol{f}(t, \ldots)$ is a non-linear function of the coefficients of the numerical solution (7.14). The standard approach would be to linearize the right-hand sides of eqns. (7.15) and to solve the resulting linear system of equations iteratively. The approach would indeed work very well. It would, however, require the knowledge of the partial derivatives of the function $\boldsymbol{f}(\ldots)$ w.r.t. all components of the solution vector and its first $n-1$ derivatives – and this might require extensive algebraic (and numerical) computations.

Fortunately, the structure of eqns. (7.15) admits a simple iterative solution. Let us assume that we managed to find an approximate solution $\boldsymbol{y}_k^I(t)$ which agrees with the true solution up to terms of order $m \geq n$ in $t - t_k$, where $n \leq m < q$:

$$\boldsymbol{y}_k^I(t) = \boldsymbol{y}_k(t) + \boldsymbol{O}\left((t - t_k)^{m+1}\right) . \tag{7.50}$$

A solution $\boldsymbol{y}_k^{I+1}(t)$ of the next higher order $m + 1$ is found by replacing the "true" solution values on the right-hand side of equations (7.15) by the approximation (7.50) and by setting the order of the approximation on the left-hand side to $m + 1$ (and not to q):

$$\sum_{l=n}^{m+1} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!} \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} = \boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k^I(t_{k_j}), \dot{\boldsymbol{y}}_k^I(t_{k_j}), \ldots, \left(\boldsymbol{y}_k^I\right)^{(n-1)}(t_{k_j})\Big) ,$$
$$j = 1, 2, \ldots, m - n + 2 . \tag{7.51}$$

The "new" condition equations (7.51) are much easier to handle than the original ones, because the right-hand sides are now known functions of time – exactly as in the case of numerical quadrature (7.49)! Elegance and simplicity have their price: For a collocation method of order q, initialized with the Euler approximation (7.11, 7.12), we have in the first step $m = n$, which implies that $q - n$ iteration steps are required to solve eqns. (7.15) "from scratch". Fortunately, this price has to be paid only once in the entire integration covering the interval $I = [t_0, t_N]$, namely in the first subinterval I_0. In subsequent subintervals the numerical solution of order q of the preceding interval may be used for initialization. This reduces the number of necessary iteration steps dramatically, normally to 1-2 steps.

The iteration process defined by eqns. (7.51) is convincingly simple and easy to implement into a subprogram. We have not yet proved, however, that this iteration process actually will converge. This is achieved by the following arguments: Assuming that the term $\boldsymbol{y}^{(n-1)}(t)$ actually occurs on the right-hand side $\boldsymbol{f}(\ldots)$ of the differential equation system (which is the worst case assumption), we conclude from a correct linearization of the problem and from the assumption (7.50):

$$\begin{gathered}\boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k^I(t_{k_j}), \dot{\boldsymbol{y}}_k^I(t_{t_j}), \ldots, \left(\boldsymbol{y}_k^I\right)^{(n-1)}(t_{k_j})\Big) = \\ \boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k(t_{k_j}), \dot{\boldsymbol{y}}_k(t_{k_j}), \ldots, \boldsymbol{y}_k^{(n-1)}(t_{k_j})\Big) + \boldsymbol{O}\left((t_{k_j} - t_k)^{m-n+2}\right) .\end{gathered} \tag{7.52}$$

This result implies, in turn, that the left-hand sides of eqn. (7.51) are correct to the same order – provided the degree of the polynomials are set to $m + 1$ on the left-hand side

$$\begin{aligned}\sum_{l=n}^{m+1} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!} \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} = & \sum_{l=n}^{m+1} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!} \left(\boldsymbol{y}_{k0}^{I}\right)^{(l)} \\ & + \boldsymbol{O}\left((t_{k_j} - t_k)^{m-n+2}\right) ,\end{aligned}$$

which proves that we have

$$\left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} = \boldsymbol{y}_{k0}^{(l)} + \boldsymbol{O}\left((t_{k_j} - t_k)^{m+2-l}\right) \tag{7.53}$$

and in particular for $l = m + 1$

$$\left(\boldsymbol{y}_{k0}^{I+1}\right)^{(m+1)} = \boldsymbol{y}_{k0}^{(m+1)} + \boldsymbol{O}\left((t_{k_j} - t_k)^{1}\right) \ , \tag{7.54}$$

which proves that the iteration process (7.51), initialized either by the Euler approximation for the first interval (accurate to order n) or by the numerical solution of order q of the preceding interval, will in general converge. Should the highest derivative $\boldsymbol{y}^{(n-1)}(t)$ not occur on the right-hand side of the differential equation system (as it is, e.g., the case of the equations of motion for the planetary system or of the three-body problem Earth-Moon-Sun), the order of the approximation could even be incremented by 2 in each iteration step. This aspect is not too important for our applications, because in general there will be thousands of subintervals, and an iteration starting "from scratch", i.e., from the Euler approximation, is required only in the first subinterval I_0. In the next subintervals the solution $\boldsymbol{y}_{k-1}(t)$ (solving the initial value at t_{k-1}) may be used for initialization.

It is important, however, that the numerical solution obtained by collocation methods is identical with the true solution except for terms of order $q + 1$ or higher in $t - t_k$ (observe that all coefficients are bound by eqn. (7.53)).

7.5.2 The Local Boundary Value Problem

When solving a local boundary value problem of type (7.6), we assume that the interval containing all boundary epochs t'_i , $i = 1, 2, \ldots, n$, can be covered by one set of approximating functions (7.14). Whether or not this is true depends on the concrete problem. Important questions can be answered in Celestial Mechanics by local boundary problems. If the interval is shorter than, let us say, a quarter of a revolution, and if the orbit has only a small eccentricity, let us say $e < 0.2$, the corresponding boundary value problem can be solved in the concise form to be developed now.

Let us assume that the boundary epochs lie in the interval I_k. According to our definition of the collocation method we simply have to replace the eqns. (7.48) defining the initial value problem by the corresponding equations for the boundary values when replacing the initial by the boundary value problem:

$$\boldsymbol{y}_k(t'_i) = \sum_{l=0}^{q} \frac{1}{l!} \, (t'_i - t_k)^l \, \boldsymbol{y}_{k0}^{(l)} = \boldsymbol{y}_i \ , \quad i = 1, 2, \ldots, n \ . \tag{7.55}$$

The coefficients $\boldsymbol{y}_{k0}^{(l)}$, $l = 0, 1, \ldots, q$, emerge as the solution of the system of algebraic equations (7.55) *and* (7.15). As opposed to the initial value problem, the two parts (7.55) and (7.15) of the condition equations cannot be solved independently from each other in the general case.

There is an important exception, however, where such a separation is possible: If the function $\boldsymbol{f}(t)$ does not depend on the solution vector (nor on its first $n-1$ time derivatives), the system (7.15) can be solved independently from the system (7.55) – as a matter of fact the solution is the same, under these conditions, as in the case of the initial value problem. The remaining coefficients $\boldsymbol{y}_{k0}^{(i)}$, $i = 0, 1, \ldots, n-1$, are then obtained by eqns. (7.55):

$$
\begin{aligned}
\boldsymbol{y}_k(t_i') &= \sum_{l=0}^{q} \frac{1}{l!}\,(t_i' - t_k)^l\,\boldsymbol{y}_{k0}^{(l)} = \boldsymbol{y}_i \\
&= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t_i' - t_k)^l\,\boldsymbol{y}_{k0}^{(l)} + \sum_{l=n}^{q} \frac{1}{l!}\,(t_i' - t_k)^l\,\boldsymbol{y}_{k0}^{(l)} \qquad (7.56)\\
&= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t_i' - t_k)^l\,\boldsymbol{y}_{k0}^{(l)} + \boldsymbol{b}_i = \boldsymbol{y}_i \ , \quad i = 1, 2, \ldots, n \ ,
\end{aligned}
$$

where

$$
\boldsymbol{b}_i \overset{\text{def}}{=} \sum_{l=n}^{q} \frac{1}{l!}\,(t_i' - t_k)^l\,\boldsymbol{y}_{k0}^{(l)} \ , \quad i = 1, 2, \ldots n \ ,
$$

are known functions after the solution of the system (7.15). Therefore, the equations (7.56) represent for each component of the solution vector a linear system of n equations for the determination of the first n coefficients of the development (7.14).

This means, on the other hand, that we may again set up a very efficient iterative solution for the combined system (7.55, 7.15), where the linear version of the system (7.15) is, as a matter of fact, identical with the corresponding system

$$
\sum_{l=n}^{m+1} \frac{(t_{ki} - t_k)^{l-n}}{(l-n)!} \left(\boldsymbol{y}_{k0}^{(I+1)}\right)^l = \boldsymbol{f}\left(t_{k_j}, \boldsymbol{y}_k^I(t_{k_j}), \dot{\boldsymbol{y}}_k^I(t_{k_j}), \ldots, \left(\boldsymbol{y}_k^I\right)^{(n-1)}(t_{k_j})\right)
$$

$$
j = 1, 2, \ldots, m+2-n \ , \qquad (7.57)
$$

when solving the initial value problem. After the solution of the linear system (7.57) the first n coefficients are obtained in analogy to eqns. (7.56) as

$$\sum_{l=0}^{n-1} \frac{1}{l!} \, (t'_i - t_k)^l \, \left(\boldsymbol{y}_{k0}^{I+1} \right)^{(l)} = \boldsymbol{y}_i \, - \, \boldsymbol{b}_i^{I+1} \; , \quad i = 1, 2, \ldots, n \; . \tag{7.58}$$

From the above developments the conclusion may be drawn that collocation methods are very flexible. They allow the solution of initial value problems and of local boundary value problems, basically with one and the same algorithm. In the case of initial value problems, the first n coefficients are defined once and for all by eqn. (7.48), whereas in the case of local boundary value problems these coefficients are obtained as a solution of the linear equations (7.58) after having solved the system (7.57) of linear condition equations for the coefficients of order $l \geq n$.

The iteration process (7.57) has to be initialized. When dealing with an initial value problem, the process was initialized with the Euler approximation (7.11, 7.12). In the general case, the same procedure cannot be applied when solving boundary value problems, because we do not know the first (and higher) derivatives of the solution vector available at any epoch t within the subinterval I_k. It is therefore necessary to initialize the iteration process with the interpolation polynomial of degree $m = n - 1$ defined by the boundary epochs and values. If the function $\boldsymbol{f} \overset{\text{def}}{=} \boldsymbol{f}(t, \boldsymbol{y})$ does not depend on the derivatives of $\boldsymbol{y}$, it is possible to initialize with a polynomial of degree $m = 2n - 1$, where the coefficients are defined by the n boundary conditions and n conditions of type (7.57) at the boundary epochs. This means that in Celestial Mechanics, if the there are no velocity-dependent forces, the initialization may be performed with the integration order $q = 3(!)$.

7.5.3 Efficient Solution of the Initial Value Problem

Collocation methods require the solution of one linear system (7.51) of condition equations in each subinterval I_k. This implies the inversion of matrices (even for every iteration step) within each subinterval I_k. Matrix inversions are time consuming operations (see, e.g., [88]). Using the same interval subdivision (see Figure 7.4) relative to the subinterval boundaries t_k and t_{k+1} for all subintervals I_k, $k = 0, 1, \ldots, N-1$, the coefficient matrices may be made identical in all intervals – provided the coefficients $\boldsymbol{y}_{k0}^{(l)}$ are scaled in an appropriate way. This will be done subsequently. Also, we will introduce a very dense and (hopefully) elegant matrix notation.

So far we only requested that all collocation epochs t_{k_j}, $j = 1, 2, \ldots, q{+}1{-}n$, are different, implying that there are many different ways how to select the epochs t_{k_j}, $i = 1, 2, \ldots, q{+}1{-}n$, within the interval I_k. One obvious (but not necessarily the best) way, leading, however, to a very transparent algorithm, is to select an equidistant subdivision covering the entire subinterval:

$$t_{k_j} \overset{\text{def}}{=} t_k + (j-1)\frac{h_k}{q-n}\ , \quad j = 1, 2, \ldots, q+1-n\ , \tag{7.59}$$

where $h_k \overset{\text{def}}{=} t_{k+1} - t_k$ was already defined as the length of subinterval I_k.

The condition equations (7.51) can now be written with an *interval-independent matrix of coefficients*:

$$\begin{aligned} \sum_{l=n}^{m+1} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!} \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} &= \sum_{l=n}^{m+1} (j-1)^{l-n} \left\{ \frac{1}{(l-n)!} \left(\frac{h_k}{q-n}\right)^{l-n} \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} \right\} \\ \overset{\text{def}}{=} \sum_{l=n}^{m+1} (j-1)^{l-n}\, \boldsymbol{c}_{k_l}^{I+1} &= \boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k^I(t_{k_j}), \dot{\boldsymbol{y}}_k^I(t_{k_j}), \ldots, \left(\boldsymbol{y}_k^I\right)^{(n-1)}(t_{k_j})\Big)\ , \\ & \qquad j = 1, 2, \ldots, q+1-n\ , \end{aligned} \tag{7.60}$$

where

$$\boldsymbol{c}_{k_l}^{I+1} \overset{\text{def}}{=} \frac{1}{(l-n)!} \left(\frac{h_k}{q-n}\right)^{l-n} \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)}\ , \quad l = n, n+1, \ldots, q\ . \tag{7.61}$$

The matrix of coefficients merely consists of powers of integer numbers. The above equations are written component by component (as mentioned we obtain one system of $q+1-n$ equations per component), and then combined into the matrix equation

$$\mathbf{M}\,\mathbf{C}_k^{I+1} = \mathbf{F}_k^I\ , \tag{7.62}$$

where

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \ldots \\ 1 & 2 & 4 & 8 & 16 & 32 & 64 & 128 & 256 & \ldots \\ 1 & 3 & 9 & 27 & 81 & 243 & 729 & 2187 & 6561 & \ldots \\ 1 & 4 & 16 & 64 & 256 & 1024 & 4096 & 16384 & 65536 & \ldots \\ 1 & 5 & 25 & 125 & 625 & 3125 & 15625 & 78125 & 390625 & \ldots \\ 1 & 6 & 36 & 216 & 1296 & 7776 & 46656 & 279936 & 167916 & \ldots \\ 1 & 7 & 49 & 343 & 2401 & 16807 & 117649 & 823543 & 5764801 & \ldots \\ 1 & 8 & 64 & 512 & 4096 & 32768 & 262144 & 2097152 & 16777216 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix} \tag{7.63}$$

and

$$\mathbf{C}_k^{I+1} = \begin{pmatrix} \left[\boldsymbol{c}_{k_n}^{I+1}\right]^T \\ \left[\boldsymbol{c}_{k_{n+1}}^{I+1}\right]^T \\ \cdots \\ \cdots \\ \cdots \\ \left[\boldsymbol{c}_{k_q}^{I+1}\right]^T \end{pmatrix} \ ; \quad \mathbf{F}_k^I = \begin{pmatrix} \left[\boldsymbol{f}^I(t_{k_1})\right]^T \\ \left[\boldsymbol{f}^I(t_{k_2})\right]^T \\ \cdots \\ \cdots \\ \cdots \\ \left[\boldsymbol{f}^I(t_{k_{q+1-n}})\right]^T \end{pmatrix} . \tag{7.64}$$

$\mathbf{C}_k^{I+1}$ and $\mathbf{F}_k^I$ are matrices with $q+1-n$ lines and d columns, where each column stands for a component of the differential equation system. If a collocation method of order q is applied to a differential equation system of order n, the dimension of the matrix $\mathbf{M}$ is $q+1-n$, and independent of the dimension d of the differential equation system. Note, as well, that the *elements of this matrix do neither depend on q nor on n.*

The solution of equation (7.62) may be written as

$$\mathbf{C}_k^{I+1} = \mathbf{M}^{-1}\,\mathbf{F}_k^I \; . \tag{7.65}$$

Solution (7.65) is elegant, but it has two important disadvantages:

1. Whereas the elements of matrix $\mathbf{M}$ are order-independent, the same is not true for the inverse matrix.
2. An algorithm derived from eqn. (7.65) cannot be recommended from the numerical point of view, because differences of big numbers (resulting in small numbers) have to be formed, what may lead to a loss of significant digits.

Both disadvantages are removed by transforming the original condition equations (7.62) into a scheme, where on the right-hand side the matrix $\mathbf{F}_k^I$ is replaced by the matrix consisting of the differences (from order zero to order $q-n$) of the original function values $\boldsymbol{f}(t_{k_j})$. Such a difference scheme, illustrated by Figure 7.8, is defined by:

$$\begin{aligned} \mathit{\Delta}_{k_j}^{[0]} &\overset{\text{def}}{=} \boldsymbol{f}(t_{k_j})\;, \quad j=1,2,\ldots \\ \mathit{\Delta}_{k_j}^{[l+1]} &\overset{\text{def}}{=} \mathit{\Delta}_{k_{j+1}}^{[l]} - \mathit{\Delta}_{k_j}^{[l]}\;, \quad j=1,2,\ldots\;,\; l=0,1,2,\ldots\;. \end{aligned} \tag{7.66}$$

The differences defined above are also called *forward differences.*

Formally, the transformation of the original system (7.62) of condition equations into one based on the forward differences is achieved by multiplying this matrix equation from the left with the auxiliary matrices $\mathbf{D}_1$, $\mathbf{D}_2$, ..., $\mathbf{D}_{q-n}$, and so on, where

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0\ldots \\ -1 & 1 & 0 & 0 & 0\ldots \\ 0 & -1 & 1 & 0 & 0\ldots \\ 0 & 0 & -1 & 1 & 0\ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix} ; \; \mathbf{D}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0\ldots \\ 0 & 1 & 0 & 0 & 0\ldots \\ 0 & -1 & 1 & 0 & 0\ldots \\ 0 & 0 & -1 & 1 & 0\ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix} ; \; \mathbf{D}_3 = \ldots\;. \tag{7.67}$$

Through this series of transformations the matrix $\mathbf{M}$ becomes an upper triangular matrix $\tilde{\mathbf{M}}$. The result is:

$$\mathbf{D}_{q-n}\,\ldots\,\mathbf{D}_2\,\mathbf{D}_1\,\mathbf{M}\,\mathbf{C}_k^{I+1} = \mathbf{D}_{q-n}\,\ldots\,\mathbf{D}_2\,\mathbf{D}_1\,\mathbf{F}_k^I \tag{7.68}$$

$$\begin{array}{lllllll}
\boldsymbol{f}(t_{k_1}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_1} & & & & & & \\
 & \Delta^{[1]}_{k_1} & & & & & \\
\boldsymbol{f}(t_{k_2}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_2} & & \Delta^{[2]}_{k_1} & & & & \\
 & \Delta^{[1]}_{k_2} & & \Delta^{[3]}_{k_1} & & & \\
\boldsymbol{f}(t_{k_3}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_3} & & \Delta^{[2]}_{k_2} & & \Delta^{[4]}_{k_1} & & \\
 & \Delta^{[1]}_{k_3} & & \Delta^{[3]}_{k_2} & & \Delta^{[5]}_{k_1} & \\
\boldsymbol{f}(t_{k_4}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_4} & & \Delta^{[2]}_{k_3} & & \Delta^{[4]}_{k_2} & & \Delta^{[6]}_{k_1} \\
 & \Delta^{[1]}_{k_4} & & \Delta^{[3]}_{k_3} & & \Delta^{[5]}_{k_2} & \\
\boldsymbol{f}(t_{k_5}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_5} & & \Delta^{[2]}_{k_4} & & \Delta^{[4]}_{k_3} & & \\
 & \Delta^{[1]}_{k_5} & & \Delta^{[3]}_{k_4} & & & \\
\boldsymbol{f}(t_{k_6}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_6} & & \Delta^{[2]}_{k_5} & & & & \\
 & \Delta^{[1]}_{k_6} & & & & & \\
\boldsymbol{f}(t_{k_7}) \stackrel{\text{def}}{=} \Delta^{[0]}_{k_7} & & & & & &
\end{array}$$

Fig. 7.8. Visualization of forward differences up to order 6

and the final result may be written as

$$\tilde{\mathbf{M}}\,\mathbf{C}_k^{I+1} = \tilde{\mathbf{F}}_k^I\ . \tag{7.69}$$

The matrix $\tilde{\mathbf{M}}$ up to order $q = n + 10$ has the form:

$$\tilde{\mathbf{M}} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 2 & 6 & 14 & 30 & 62 & 126 & 254 & 510 & 1022 \\
0 & 0 & 0 & 6 & 36 & 150 & 540 & 1806 & 5796 & 18150 & 55980 \\
0 & 0 & 0 & 0 & 24 & 240 & 1560 & 8400 & 40824 & 186480 & 818520 \\
0 & 0 & 0 & 0 & 0 & 120 & 1800 & 16800 & 126000 & 834120 & 5103000 \\
0 & 0 & 0 & 0 & 0 & 0 & 720 & 15120 & 191520 & 1905120 & 16435440 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 5040 & 141120 & 2328480 & 29635200 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 40320 & 1451520 & 30240000 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 362880 & 16329600 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3628800
\end{pmatrix} . \tag{7.70}$$

The right-hand side of eqns. (7.69) contain the following differences of the original function values $\boldsymbol{f}^I(t_{k_j})$:

$$\tilde{\mathbf{F}}_k^I = \begin{pmatrix} \left[\boldsymbol{f}^I(t_{k_1})\right]^T \\ \left[\Delta_{k_1}^{[1]}\right]^T \\ \left[\Delta_{k_1}^{[2]}\right]^T \\ \dots \\ \dots \\ \dots \\ \left[\Delta_{k_1}^{[q-n]}\right]^T \end{pmatrix} , \tag{7.71}$$

where the difference scheme has to be formed according to the pattern (7.66). The differences used are the first ones given in each column in Figure 7.8.

The solution of eqn. (7.69) may now be written as

$$\mathbf{C}_k^{I+1} = \tilde{\mathbf{M}}^{-1}\,\tilde{\mathbf{F}}_k^I \ , \tag{7.72}$$

where up to order $q = n + 10$:

$$\tilde{\mathbf{M}}^{-1} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -\frac{1}{2} & \frac{1}{3} & -\frac{1}{4} & \frac{1}{5} & -\frac{1}{6} & \frac{1}{7} & -\frac{1}{8} & \frac{1}{9} & -\frac{1}{10} \\
0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{11}{24} & -\frac{5}{12} & \frac{137}{360} & -\frac{7}{20} & \frac{363}{1120} & -\frac{761}{2520} & \frac{7129}{25200} \\
0 & 0 & 0 & \frac{1}{6} & -\frac{1}{4} & \frac{7}{24} & -\frac{5}{16} & \frac{29}{90} & -\frac{469}{1440} & \frac{29531}{90720} & -\frac{1303}{4032} \\
0 & 0 & 0 & 0 & \frac{1}{24} & -\frac{1}{12} & \frac{17}{144} & -\frac{7}{48} & \frac{967}{5760} & -\frac{89}{480} & \frac{4523}{22680} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{120} & -\frac{1}{48} & \frac{5}{144} & -\frac{7}{144} & \frac{1069}{17280} & -\frac{19}{256} \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{720} & -\frac{1}{240} & \frac{23}{2880} & -\frac{1}{80} & \frac{3013}{172800} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5040} & -\frac{1}{1440} & \frac{13}{8640} & -\frac{1}{384} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{40320} & -\frac{1}{10080} & \frac{29}{120960} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{362880} & -\frac{1}{80640} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3628800}
\end{pmatrix} . \tag{7.73}$$

Note that the elements of matrix $\tilde{\mathbf{M}}^{-1}$, as the inverse of an upper triangular matrix, do not change, if the order q of the approximation changes. Therefore it is only necessary to store *one* (triangular) matrix of maximum order $q_{\max}$ for all algorithms with orders $q \leq q_{\max}$. This removes the first of the disadvantages mentioned after eqn. (7.65). The second disadvantage has been removed also by introducing the differences of the elements of matrix $\mathbf{F}_k^I$ (and by not using the matrix elements themselves). Note also, that the elements of the coefficient matrix $\tilde{\mathbf{M}}^{-1}$ are given as fractions of integers. The matrix was generated with a computer program.

The diagonal of matrix $\tilde{\mathbf{M}}^{-1}$ obviously contains the terms $1/l!$, $l = 1, 2, \dots,$ $q{-}n$. As the part of the matrix below the diagonal only contains zero elements and as the terms $\boldsymbol{c}_{k_l}$ are defined by eqn. (7.61), we may conclude that

$$\left(\boldsymbol{y}_{k0}^{I+1}\right)^{(q)} = \left(\frac{q-n}{h_k}\right)^{q-n} \mathit{\Delta}_{k_1}^{[q-n]} + \boldsymbol{O}(h_k) \ . \tag{7.74}$$

The above equation says that the $(q-n)$-th forward difference of the forces may be interpreted (in a modest approximation and apart from a scaling factor) as the q-th derivative of the solution vector $\boldsymbol{y}_k(t)$ at t_k. The result will be used when implementing automatic stepsize control.

Equation (7.72) tells that the coefficients $\boldsymbol{c}_{k_l}$, $l = n, n+1, \ldots, q$, are linear combinations of the forward differences $\mathit{\Delta}_{k_1}^{[l]}$, $l = 0, 1, \ldots, q-n$. Consequently (see eqn. (7.73)), the solution vector $\boldsymbol{y}_k(t)$ may be represented as a linear combination of the same differences. Let us give the resulting linear combination in explicit form:

$$\begin{aligned}
\boldsymbol{y}_k^{I+1}(t) &= \sum_{l=0}^{q} \frac{1}{l!}\,(t-t_k)^l \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} \\
&= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t-t_k)^l\, \boldsymbol{y}_{k0}^{(l)} + \sum_{l=n}^{q} \frac{1}{l!}\,(t-t_k)^l \left(\boldsymbol{y}_{k0}^{I+1}\right)^{(l)} \\
&= \ldots + \sum_{l=n}^{q} \frac{1}{l!}\,(t-t_k)^l\,(l-n)! \left(\frac{q-n}{h_k}\right)^{l-n} \boldsymbol{c}_{k_l}^{I+1} \\
&= \ldots + \sum_{l=n}^{q}\sum_{j=l}^{q} \frac{1}{l!}\,(t-t_k)^l\,(l-n)! \left(\frac{q-n}{h_k}\right)^{l-n} \tilde{\mathrm{M}}_{l-n+1,j-n+1}^{-1}\, \mathit{\Delta}_{k_1}^{[j-n]}
\end{aligned} \tag{7.75}$$

Introducing the interval-independent time argument by

$$\tau = \left(\frac{q-n}{h_k}\right)\,\left(t-t_k\right) \ , \tag{7.76}$$

we may further develop the above expression to obtain eventually:

$$\begin{aligned}
\boldsymbol{y}_k(t) &= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t-t_k)^l\, \boldsymbol{y}_{k0}^{(l)} \\
&+ \sum_{j=n}^{q} \left\{ \sum_{l=n}^{j} \frac{(l-n)!}{l!}\, \tilde{\mathrm{M}}_{l-n+1,j-n+1}^{-1}\, \tau^l \right\} \left(\frac{h_k}{q-n}\right)^{n} \mathit{\Delta}_{k_1}^{[j-n]} \ .
\end{aligned} \tag{7.77}$$

Observe that the term in brackets $\{\ldots\}$ does not contain any terms depending on the subinterval I_k . For the purpose of error control we do not only need the formula for the solution vector, but also for its first $n-1$ derivatives. Taking into account that

$$\frac{d^i}{dt^i} = \left(\frac{q-n}{h_k}\right)^i \frac{d^i}{d\tau^i} \ , \tag{7.78}$$

we obtain the result:

$$
\begin{aligned}
\left(\boldsymbol{y}_k^{I+1}\right)^{(i)}(t) = {} & \sum_{l=i}^{n-1} \frac{1}{(l-i)!}\,(t-t_k)^{l-i}\,\boldsymbol{y}_{k0}^{(l)} \\
& + \sum_{j=n}^{q}\left\{\sum_{l=n}^{j}\frac{(l-n)!}{(l-i)!}\,\tilde{\mathrm{M}}^{-1}_{l-n+1,j-n+1}\,\tau^{l-i}\right\}\left(\frac{h_k}{q-n}\right)^{n-i}\Delta_{k_1}^{[j-n]}\ , \\
& \qquad\qquad\qquad\qquad i = 0, 1, \ldots, n-1\ .
\end{aligned}
\tag{7.79}
$$

The collocation method implemented in the program system contains the coefficient matrices up to degree and order $q - n = 14$. The maximum order thus would be $q = 16$ for a differential equation system of order $n = 2$. The algorithm supplied in the program system is able to handle any order $n \geq 1$ of the differential equation system. It may be initialized either by the Euler approximation or by the numerical solution of a previous step.

7.5.4 Integrating a Two-Body Orbit with a High-Order Collocation Method: An Example

The power of a high-order collocation method (and of standard computer equipment like PCs, notebooks, etc.) is illustrated by integrating the orbit of a virtual minor planet with elements given by Table 7.1 over a time period of one thousand years with program NUMINT (program PLASYS might be used for the same purpose).

The processing characteristics are as follows:

- A 12th order collocation method with constant stepsize was used.
- Two program runs were made with stepsizes of $h = 30$ days and $h = 100$ days.
- After initialization, one iteration step was performed per sub-interval I_k.
- One output record (containing the osculating elements and the radial, along-track, and out-of-plane coordinate errors) was stored per 100 days.

The general program output, containing statistical information of the program run, is reproduced (partially) in Figure 7.9 for the program runs with 100 days stepsize. (The initial osculating elements and some additional information were excluded.) Figure 7.9 tells that the program run only lasted for 0.4 s on a PC with a 1.4 GHz processor. Processing times are of course hardware-dependent. The example shows, however, that test runs over tens of thousands of revolutions do not pose serious problems, today. In the 1960s such tests had the tendency to last for hours, even on mainframe computers. The average number of evaluations of the right-hand sides of the differential

```
NUMERICAL SOLUTION OF ORDINARY  DEQ-SYSTEMS DATE: 21-SEP-02    TIME: 06:56
***************************************************************************
METHOD OF SOLUTION : COLLOCATION OF ORDER 12
******************
STEP SIZE          :     100.000000 DAYS
# STEPS            :           3653
# COMP (RHS)/REV   :           304.
CPU (TOT)          :           0.40 S
MAX ERROR  IN A    :     0.2834D-10 AU
MAX ERROR  IN I    :     0.2560D-12 DEG
MAX ERROR  IN ANOM.:     0.3440D-07 AU
MAX ERROR  IN POS. :     0.3440D-07 AU
ADDITIONAL INFORMATION FOR COLLOCATION METHODS:
**********************************************
ORDER OF DEQ-SYSTEM                             :        2
INITIALIZE AT EACH STEP                         :       NO
INCREMENT ORDER OF APPROX PER ITERATION STEP    :        2
NUMBER OF ITERATION STEPS AFTER INITIAL STEP    :        1
```

Fig. 7.9. Output of program NUMINT for run with stepsize of 100 days

equation system per revolution is a computer-independent quantity. Roughly, one function evaluation was required per degree swept by the radius vector in the test illustrated in Figure 7.9.

Figure 7.10 shows the errors in the semi-major axis a (top row), the eccentricity e (second row), the longitude of perihelion ω (third row), the mean anomaly σ_0 at time t_0 (fourth row), and the mean anomaly difference $\delta\sigma \overset{\text{def}}{=} n(t)\big(t - T_0(t)\big) - n_0\big(t - T_0(t_0)\big)$ at time t. The left-hand column of the figure corresponds to the stepsize of $h = 100$ days, the right-hand column to that of $h = 30$ days. The units are 10^{-10} AU (first row), 10^{-10} (second row), 10^{-10} degrees in rows $3-5$. It does not make sense to document the errors of the inclination i and of the longitude of the node Ω, because these values are in essence “error-free” when the two-body problem is integrated (all vectors are linear combinations of the initial position- and velocity-vector).

All errors in Figure 7.10 are small. The errors on the right-hand side are, however, smaller by a factor of about 1000 than those in the corresponding figures on the left-hand side. As the ratio of the stepsizes is 0.3, one would, however, expect the ratio of the errors to be of the order of $0.3^{13} \approx 1.6 \cdot 10^{-7}$. The expectations clearly are *not* met in Figure 7.10. The “failure” is explained by the circumstance that two different kinds of errors are visible in Figure 7.10: on the left-hand sides we see the accumulated errors due to the truncation of the numerical solution, subsequently called *approximation errors*, on the right-hand side we see the accumulated *rounding errors*, due to the calculation with a finite number of significant digits. The accumulation of the two error types will be further discussed in section 7.7. The accumulated approximation errors clearly show a systematic behavior, the accumulated rounding errors show a random pattern.

There is a distinct difference in behavior between the accumulated approximation errors (left column in Figure 7.10) in the first three and the last two

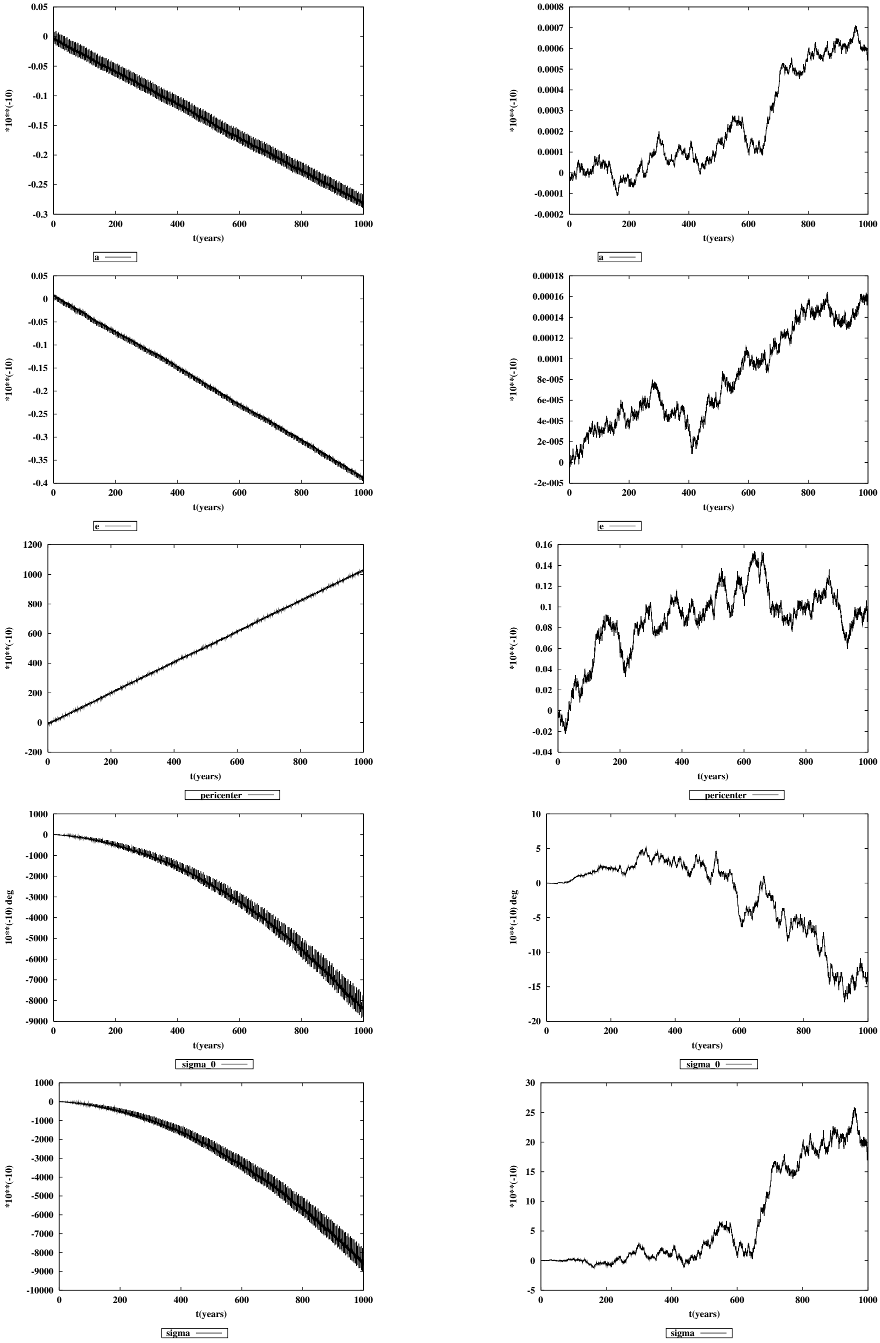


Fig. 7.10. Accumulation of errors in semi-major axis a, eccentricity e, longitude of perihelion ω, mean anomalies σ_0, and $\delta\sigma$ over 1000 years (left: stepsize=100 days, right: stepsize=30 days)

rows: a linear trend is observed in the former, a quadratic growth in the latter case. The error in $\delta\sigma(t)$ may be explained (approximately) by the linear decrease of the semi-major axis:

$$\delta\sigma(t) = \int\limits_{t_0}^{t} \delta n(t')\ dt' \approx -\frac{3}{2}\frac{n_0}{a_0} \int\limits_{t_0}^{t} \delta a(t')\ dt'\ . \tag{7.80}$$

From Figure 7.10 (top, left) we may extract

$$\delta a(t) \approx -0.27 \cdot 10^{-10} \left(\tilde{t} - \tilde{t}_0\right) \text{ AU}\ , \tag{7.81}$$

where the time argument $\tilde{t} - \tilde{t}_0$ is measured in units of 1000 years relative to the initial epoch t_0. Using this formula in eqn. (7.80) gives the following relation for the growth of $\delta\sigma(t)$:

$$\delta\sigma(t) = \int\limits_{t_0}^{t} \delta n(t')\ dt' \approx \frac{3}{4}\frac{n_0}{a_0} \cdot 0.27 \cdot 10^{-10}\ \left(\tilde{t} - \tilde{t}_0\right)^2 = 7366 \cdot 10^{-10}\ \left(\tilde{t} - \tilde{t}_0\right)^2\ , \tag{7.82}$$

where n_0 has to be measured in degrees per 1000 years. This relation is confirmed approximately in Figure 7.10 (bottom, left).

Often, one is not particularly interested in the integration error associated with the orbital elements (or in the quantity $\sigma(t)$), but rather in the errors of the rectangular coordinates of the trajectory. In such cases it is wise *not* to display the errors in the coordinates of the inertial system, but rather (a) in radial, (b) along-track, and (c) in the out-of-plane direction. Figure 7.11 illustrates this error decomposition. As expected from the error accumulation in the mean anomaly $\sigma(t)$, the error in the along-track direction dominates the errors in the other two directions. It is easy to interpret the along-track errors by re-scaling the results obtained for $\delta\sigma(t)$. In view of the remarks made concerning the elements i and Ω, it cannot amaze that the errors in the out-of-plane direction are very small – negligible compared to the other two components. The errors in the radial direction are a consequence of the error in $\sigma(t)$, as well. The polar equation (4.63) for the ellipse, expressed by the eccentric anomaly E as argument, may be approximated as follows:

$$r = a\,(1 - e\ \cos E) \approx a\,(1 - e\ \cos\sigma)\ . \tag{7.83}$$

Consequently, an error in σ induces a periodic error in radial direction:

$$\delta r(t) \approx a\,e\,\delta\sigma(t)\ \sin\sigma\ . \tag{7.84}$$

This explains the pattern of the perturbations in radial direction in Figure 7.11.

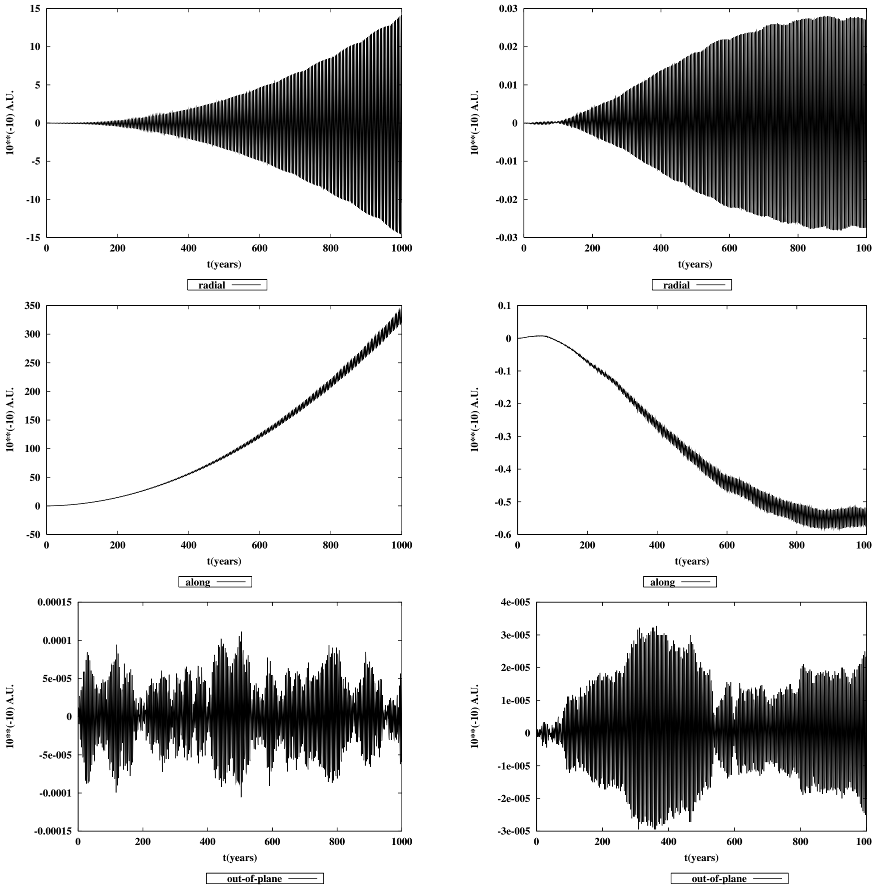


Fig. 7.11. Accumulation of errors in radial (top row), along-track (second row), and out-of-plane directions (bottom row) (left: stepsize=100 days, right: stepsize=30 days)

7.5.5 Local Error Control with Collocation Algorithms

Motivation. The numerical experiments in the preceding section might have given the impression that collocation methods with a constant stepsize (constant length $h_k = h$ of subintervals I_k) are sufficient to solve all kinds of equations of motion. The situation changes considerably, however, as soon as the orbital eccentricity e is growing. Figure 7.12 demonstrates what may happen, if a well established procedure *without* error control is used to compute an orbit with an eccentricity of $e = 0.9$ (the other orbital elements were kept to the values of Table 7.1).

Despite the fact that a constant stepsize of only $h = 6$ days was used, the result is discouraging. Even the errors in the semi-major axis assume values of a fraction of one percent – compared to the results obtained in Figure 7.10 with a stepsize which was five or even approximately 15 times larger, really not a good achievement. The results in Figure 7.12 indicate, however, that the collocation method with constant stepsize, which proved to be quite efficient for orbits with small eccentricities, still might be used successfully for orbits with large eccentricities – provided the stepsize is further reduced. As a matter of fact, acceptable results are obtained with stepsizes of the order of $h = 0.5 - 1$ days. The price to achieve results of acceptable quality is high, though. The computational costs increase by a factor of about $30 - 100$ when integrating the orbit of an eccentricity of $e \approx 0.9$ compared to one of $e \approx 0.1$. This poor efficiency is a consequence of the fact that the stepsize has to be adapted to the worst case, i.e., to the motion near perihelion.

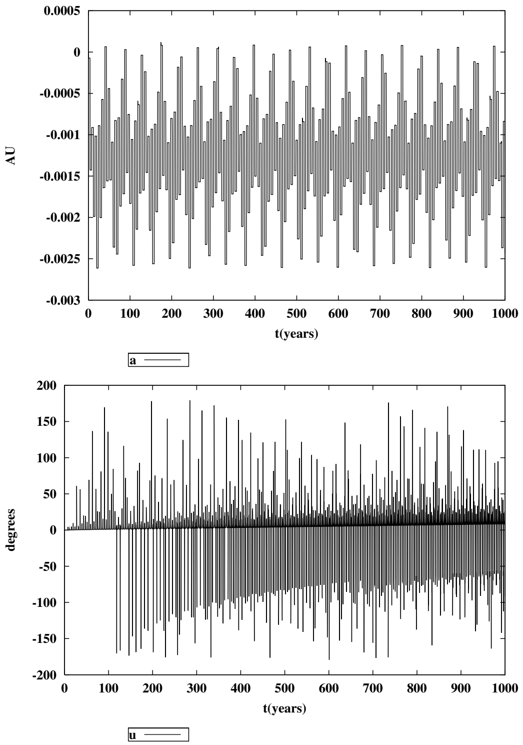


Fig. 7.12. Errors in semi-major axis a and argument of latitude u over 1000 years for an orbit with $e = 0.9$ using stepsize $h = 6$ days

Principles of Stepsize Control. The computational situation encountered above could be significantly improved, if accurate information concerning the terms neglected in the numerical solution would be available in each interval I_k. We might then adapt the stepsize with the goal to keep the error in the position and/or the velocity associated with the initial value problem of interval I_k below a specified (user-defined) limit. We have seen that the numerical and the true solutions of the initial value problem in the subinterval I_k differ only by terms of the order $\boldsymbol{O}(t-t_k)^{q+1}$. Unfortunately there is no way of estimating these terms (except by actually increasing the order internally to (at least) $q+1$). Therefore, we have to assess the error of the collocation step with the highest forward difference of $\boldsymbol{f}(t)$ of order $q-n$, which is still taken into account. Using eqn. (7.79) to represent the state vector and eqn. (7.74) to replace the differences of order $q-n$ of the functions $\boldsymbol{f}(t_{k_j})$ by the derivative of order q of the solution vector, we obtain the approximation error in the subinterval I_k by the last term of order $j=q$ included in eqn. (7.79):

$$\begin{aligned}\boldsymbol{\varepsilon}_k^{(i)}(t) &\overset{\text{def}}{=} \left\{\sum_{l=n}^{q} \frac{(l-n)!}{(l-i)!}\, \tilde{\mathrm{M}}^{-1}_{l+1-n,q+1-n}\, \tau^{l-i}\right\} \left(\frac{h_k}{q-n}\right)^{q-i} \boldsymbol{y}_{k0}^{(q)} \\ &\overset{\text{def}}{=} e_r^{(i)}(\tau) \left(\frac{h_k}{q-n}\right)^{q-i} \boldsymbol{y}_{k0}^{(q)}\ , \quad i=0,1,\ldots,n-1\ ,\end{aligned} \tag{7.85}$$

where the scalar function $e_r(\tau)$ describes the (relative) propagation of the errors in the subinterval I_k. Observe that $e_r(\tau)$ does not depend on the specific interval I_k. The relative time argument τ is defined by eqn. (7.76). If we accept the above crude (and usually pessimistic) approximation of the local error, we obtain the optimum length $h_{k,\mathrm{opt}}$ of the interval I_k rather easily, and even in “in real time”.

For this purpose, the formula (7.85) is evaluated for the time argument $t = t_k + h_k = t_{k+1}$ of the new initial values and for the derivative of order $i = n-1$. We are thus controlling the error of the $(n-1)$-th derivative of the solution vector. We can compare the actual error $\boldsymbol{\varepsilon}_k^{(n-1)}(t_{k+1})$, as emerging from formula (7.85), component by component, with the externally provided tolerance $\tilde{\varepsilon}$. Let

$$\left|\, y_{k0,i_{\max}}^{(n-1)} \right| \overset{\text{def}}{=} \max\left\{ \left| y_{k0,1}^{(n-1)} \right|, \left| y_{k0,2}^{(n-1)} \right|, \ldots, \left| y_{k0,d}^{(n-1)} \right| \right\} \tag{7.86}$$

be the component of maximum absolute value of the vector $\boldsymbol{y}_{k0}^{(n-1)}$.

The optimum stepsize $h_{k,\mathrm{opt}}$ is obtained by equating the error function $\boldsymbol{\varepsilon}_k^{(n-1)}(t_{k+1})$ for the component $i_{\max}$ to the maximally tolerated error $\tilde{\varepsilon}$ (observe that $\tau(t_{k+1}) = q-n$):

$$\left| e_r^{(n-1)}(q-n) \right| \left(\frac{h_{k,\mathrm{opt}}}{q-n}\right)^{q+1-n} \left| y_{k0,i_{\max}}^{(q)} \right| = \tilde{\varepsilon}\ , \tag{7.87}$$

where we took into account that $\tau(t_{k+1}) = q - n$ for a collocation method with equidistant spacing of the collocation epochs. The above equation is solved by

$$h_{k,\mathrm{opt}} = (q-n) \left[\frac{\tilde{\varepsilon}}{\left| e_r^{(n-1)}(q-n)\, y_{k0,i_{\max}}^{(q)} \right|} \right]^{\frac{1}{q+1-n}} . \tag{7.88}$$

The criterion (7.88) asks for some comments:

- The criterion is independent of the specific problem. It may not only be used in Celestial Mechanics but for all kinds of applications.
- The criterion might be generalized by specifying not one general tolerance $\tilde{\varepsilon}$, but one per component of the $(n-1)$st derivative of the solution vector.
- The criterion might be further refined to make it depend on time t. Such criteria might be useful in perturbation problems, if an approximate solution is available.
- One might have the idea to control not only the $(n-1)$-st derivative of the solution vector, but all derivatives from $i = 0$ up to $i = n-1$. The $(n-1)$-st derivatives, however, have the lowest order of the error in h_k, which is why primarily these quantities should be kept under control.
- Criterion (7.88) is by no means the only possible way to control the local error. One might, e.g., define a limit for the relative error in the components of the solution vector. Error criteria of this type make the attempt to control the number of significant digits of a solution. For quasi-periodic solutions such criteria are usually not considered.
- Special problems may allow for better criteria.

A Case Study. Program NUMINT was used to integrate an orbit with eccentricity $e = 0.9$ – all the other orbital elements were kept to the values of Table 7.1. Figure 7.13 shows the essential part of the output statistics for this program run. Observe that the stepsize indeed varied substantially during the integration. Criterion (7.88) was used with the tolerance $\tilde{\varepsilon} = 1\cdot 10^{-13}$ AU/day.

The smallest stepsize is of the order of half a day, the longest of the order of 114 days (i.e., of the same order as the constant stepsizes used to compute the orbits of small eccentricities). Compared to the integration of orbits with small eccentricities, the computational burden was rising by a factor of about two (compared to the run with a stepsize of 30 days). This performance should be compared to the (only possible) alternative to cover the entire interval (without error control) with the constant stepsize of 0.5 days, which would have resulted in about 55000 function calls per revolution. Seen from that perspective, we have saved a factor of about 20 in processing time by using a method with an automatic stepsize control. Figure 7.14 shows the

```
NUMERICAL SOLUTION OF ORDINARY  DEQ-SYSTEMS DATE: 22-SEP-02    TIME: 05:08
***************************************************************************
PROBLEME RESTREINT  IN PLANETARY SYSTEM
***************************************
..
SEMI-MAJOR AXIS     :         2.50230307 AU
ECCENTRICITY        :         0.90000000
..
NUMBER OF REVS      :          252.6
REVOLUTION PERIOD   :         1445.8 DAYS

METHOD OF SOLUTION : COLLOCATION OF ORDER 12
******************
MIN. STEP SIZE      :      0.670557 DAYS
MAX. STEP SIZE      :    113.597915 DAYS
CPU (TOT)           :          1.92 S
# STEPS             :         24703
# COMP (RHS)/REV    :         2207.
..
ERROR TOL. IN VEL. :     0.1000D-12 AU/DAY
..
```

Fig. 7.13. Program output for run with automatic stepsize control

stepsize $h_k \overset{\text{def}}{=} h(t_k)$ as a function of time for the first 100 years of the arc. As expected, $h(t)$ is a periodic function of time and assumes its maximum values at aphelion, its minimum values at perihelion.

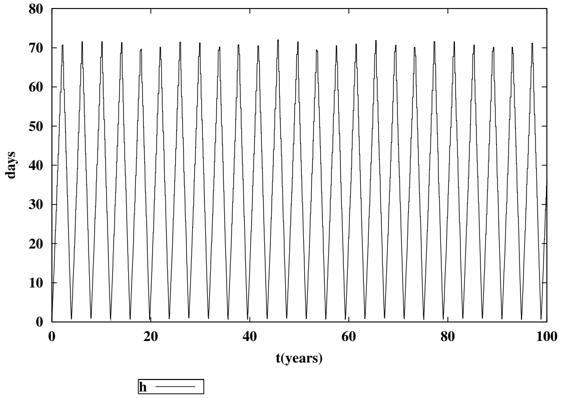


Fig. 7.14. Automatic stepsize selection in days for first 100 years

Figure 7.15 gives an impression of the quality of the results achieved, which is orders of magnitude better than that in the case documented by Figure 7.12. In this sense we may consider the experiment a success. Note in particular that the errors in the semi-major axis are multiplied by a factor of 10^{10}. The errors in a are stochastic in nature and of the order of a few 10^{-13} AU, i.e., roughly comparable to what was achieved in the case of the orbits with small eccentricities. The result is different for the argument of latitude u: We

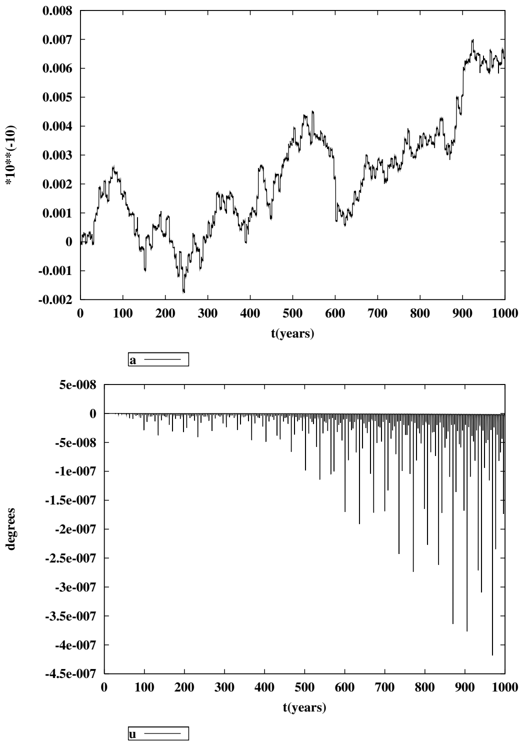


Fig. 7.15. Errors (AU) in semi-major axis a (upper figure) and argument of latitude u (lower figure) over 1000 years for an orbit with $e = 0.9$, using automatic stepsize control

observe spikes of considerable amplitudes, an effect asking for an explanation. A closer inspection shows that the performance is comparable to the case of small eccentricities everywhere except near the perihelion where the spikes in Figure 7.15 actually occur. These spikes are of the order of a few units in 10^{-7} degrees.

Is this an unacceptable error? The answer is a clear "no". The performance is typical for orbits of a large eccentricity. The characteristics of the error in the argument of latitude u in Figure 7.15 may be explained as a consequence of the development of the semi-major axis a in the same Figure. Exactly as in the case of orbits with a small eccentricity (compare eqn. (7.80)) the error induced by a into the mean anomaly $\sigma(t)$ is calculated by

$$\delta\sigma(t) \approx \int_{t_0}^{t} \delta n(t')\, dt' = -\frac{3\,n}{2\,a} \int_{t_0}^{t} \delta a(t')\, dt' \ , \tag{7.89}$$

but we are no longer allowed to identify the error in $\sigma(t)$ with the error in the argument of latitude. Using the approximation

$$\delta a(\tilde{t}) \approx 5 \cdot 10^{-16}\, \tilde{t} \text{ [AU] }, \tag{7.90}$$

where $\tilde{t}$ represents the time in years, to model the time development of $\delta a(\tilde{t})$, we obtain according to the formula (7.89)

$$\delta\sigma(\tilde{t}) = -\frac{3\,n}{4\,a}\, 5 \cdot 10^{-10} \left(\frac{\tilde{t}}{1000}\right)^2 \text{ [rad] }. \tag{7.91}$$

This equation gives $\delta\sigma(\tilde{t})$ in radian as a function of time $\tilde{t}$ in units of years. Equation (4.35) for the argument of latitude allows it to transform the error in the mean anomaly into one in the argument of latitude u:

$$\dot{u} = \frac{h}{r^2} = \sqrt{\frac{\mu}{p^3}}\, (1 + e \cos v)^2 \ , \tag{7.92}$$

from where we may conclude that in perihelion

$$\delta u(v = 0) = \sqrt{\frac{1+e}{1-e}}\, \frac{\delta\sigma}{1-e} \approx 43.6\ \delta\sigma \ , \tag{7.93}$$

where the latter value results for the eccentricity of $e = 0.9$. Near the perihelion, an error in mean anomaly translates into an error in true anomaly (and the argument of latitude) with a magnification factor of about 44 (the factor varies as a function of the eccentricity e according to the above formula). For the concrete example shown in Figure 7.15 we obtain (observe that in formula (7.91) the mean motion is needed in radian/year):

$$\delta u(v = 0) = 43.6\, \frac{180}{\pi}\, \delta\sigma = -\,6 \cdot 10^{-7} \left(\frac{\tilde{t}}{1000}\right)^2 \ [\,^\circ\,] \ . \tag{7.94}$$

This result is confirmed by Figure 7.15. It proves that the result of our integration was as good as could be expected. It shows also that for orbits with big eccentricites it is extremely difficult to control the errors in the true anomaly near perihelion. The above developments indicate that numerical tests merely giving the accuracy of the coordinates at specific epochs (e.g., always near the aphelion) might not give a full picture of the error characteristics.

The Impact of Rounding Errors. With automatic stepsize control we try to control the approximation error, not the rounding error. Formula (7.85) in essence promises that the error function obeys a h^{q-i}-law for the derivative no i of the solution, implying a dramatic reduction of errors when reducing the stepsize. This is only true, however, if we are actually capable of determining the highest derivative $\boldsymbol{y}_{k0}^{(q)}$ with reasonable accuracy (with a few significant digits) from the algorithm itself. These derivatives follow from the forward differences $\Delta_{k_1}^{[q-n]}$. The differences are in turn calculated from the function values $\boldsymbol{f}(t_{k_j})$ using the algorithm (7.66).

Rounding errors do occur in the components of the vectors $\boldsymbol{f}(t_{k_j})$. In order to understand their impact on the difference vector $\Delta_{k_1}^{[q-n]}$ of order $q-n$, it is important to review the definition of the forward differences (7.66), as illustrated by Figure 7.8. Every element in this scheme (with the exception of the elements in the first column) is the difference of the lower minus the upper element in the column to the left of the considered element.

In order to assess the order of magnitude of the error in the differences of order $q-n$ we consider one particular component and assume that (the particular component of) the vectors $\boldsymbol{f}(t_{k_j})$ (considered) with even subscripts j in Figure 7.8 are affected by a rounding error of $+\rho$, all others with one of $-\rho$. This is admittedly a rather special situation. As one may assume that the number of significant digits in the differences is smaller than or equal to the number of digits in the (considered component of the) accelerations $\boldsymbol{f}(t_{k_j})$, the difference formation process itself may be considered as free of rounding errors. This is why only the propagation of the original rounding errors (in the components) of the accelerations $\boldsymbol{f}_{k_j}$ over the scheme 7.8 has to be studied.

Figure 7.16 shows that the (absolute values of the) errors increase by a factor of 2 with each order of the difference. This power-law of error propagation holds in the general case, although the real situation is more complicated: the rounding errors of the elements of $\boldsymbol{f}(t_{k_j})$ are randomly distributed in an interval of the length of one unit of the least significant mantissa digit, centered at the true value of the $\boldsymbol{f}(t_{k_j})$, but statistically speaking, the 2^{q-n} law holds.

$$\begin{array}{lllllll}
-\rho & & & & & & \\
 & +2\,\rho & & & & & \\
+\rho & & +2^2\rho & & & & \\
 & -2\,\rho & & +2^3\rho & & & \\
-\rho & & -2^2\rho & & +2^4\,\rho & & \\
 & +2\,\rho & & -2^3\rho & & +2^5\rho & \\
+\rho & & +2^2\rho & & -2^4\rho & & 2^6\rho \\
 & -2\,\rho & & +2^3\rho & & -2^5\rho & \\
-\rho & & -2^2\rho & & +2^4\rho & & \\
 & +2\,\rho & & -2^3\rho & & & \\
+\rho & & +2^2\rho & & & & \\
 & -2\,\rho & & & & & \\
-\rho & & & & & &
\end{array}$$

Fig. 7.16. Propagation of rounding errors in a difference scheme

Figure 7.16 shows that, independently of the actual interval length, the induced error $\rho^{[q-n]}$ in one of the elements of vector $\varDelta_{k_1}^{[q-n]}$ will be of the order

$$\rho^{[q-n]} \approx 2^{q-n} \rho \ . \tag{7.95}$$

By reducing the stepsize further and further, i.e., for $h_k \rightarrow 0$, the highest differences will be fully dominated by the rounding error and will no longer contain any information concerning the solution vector. Statistically speaking we have

$$E(\varDelta_{k_1}) = \text{const.} \quad \text{for} \quad h_k \rightarrow 0 \ . \tag{7.96}$$

Equation (7.85) and (7.74), or, more directly eqn. (7.79), tell that in this case the error function for the velocity components is a linear function of h_k (and of $\varDelta_{k_1}^{[q-n]}$). This in turn implies that the optimum stepsize calculated with criterion (7.88) will be a linear function of the external tolerance $\tilde{\varepsilon}$ for $h_k \rightarrow 0$.

Table 7.2 illustrates the impact of the rounding errors. A minor planet with the elements of Table 7.1, but with $e = 0$ (implying that the stepsize should stay roughly constant for a particular value $\tilde{\varepsilon}$), was integrated over a time interval of 10000 years with the collocation method of order $q = 12$ with different error criteria. Program PLASYS was used for this purpose.

Table 7.2. Stepsize h_k as a function of the tolerance $\tilde{\varepsilon}$

$\tilde{\varepsilon}$	h_k(integration)	h_k(theory)
[AU/day]	[days]	[days]
$1 \cdot 10^{-13}$	253.5	253.5
$1 \cdot 10^{-14}$	210.9	209.2
$1 \cdot 10^{-15}$	171.2	172.7
$1 \cdot 10^{-16}$	139.0	142.6
$1 \cdot 10^{-17}$	107.6	117.7
$1 \cdot 10^{-18}$	15.05	97.1
$5 \cdot 10^{-19}$	7.7	91.7
$4 \cdot 10^{-19}$	6.2	90.0
$3 \cdot 10^{-19}$	4.7	87.8
$2 \cdot 10^{-19}$	3.2	84.9
$1 \cdot 10^{-19}$	1.6	80.2

Table 7.2 shows in column 2 the average stepsize selected by the program for different (user defined) tolerances $\tilde{\varepsilon}$. Whereas the actual stepsize obeys the power law underlying the selection criterion for the upper part of the Table (down to the value $\tilde{\varepsilon} \approx 1 \cdot 10^{-17}$ AU/day), the stepsize breaks down rather rapidly afterwards, and eventually becomes a linear function of τ, as predicted. The third column illustrates the expected

$$h(\tilde{\varepsilon}) = h\big(1 \cdot 10^{-13}\big) \cdot \left(\frac{\tilde{\varepsilon}}{1 \cdot 10^{-13}}\right)^{\frac{1}{12}}$$

law for the stepsize.

There is an important message in Table 7.2: If one tries to strive for very high accuracies using a small tolerance, the automatically selected stepsizes become very small and the efficiency is decreased instead of increased (as one should expect from automatic stepsize control). In the concrete example, the best performance is expected for values $1 \cdot 10^{-17} < \tilde{\varepsilon} < 1 \cdot 10^{-14}$.

If the limitations due to rounding errors are observed, automatic stepsize control is an excellent and very efficient tool in numerical orbit computation. It is an absolute requirement when orbits with large eccentricities are integrated. If the orbital evolution of resonant minor planets is studied, the eccentricities may vary between broad limits. Any schemes relying (essentially) on similar orbit characteristics over a long period of time might lead to unpredictable results.

In principle it is possible to develop stepsize control mechanisms taking the rounding errors into account. It would, e.g., be possible to check, whether the absolute values of the few highest differences are governed by rounding errors. Should this be the case, the criterion (7.88) should be replaced by a criterion slightly increasing the stepsize. Such advanced techniques are out of the scope of this book. We refer to [45] and [108] for further reading, to [88] for a useful and entertaining general discussion.

7.5.6 Multistep Methods as Special Collocation Methods

In section 7.4.2 it was shown that multistep methods are in principle special cases of collocation methods. This section is devoted to the development of concrete multistep algorithms. The approximating function of the initial value problem in the overlapping subintervals I_k – defined in section 7.4.2 *either* by eqns. (7.17) (for extrapolation methods) *or* by eqns. (7.18) (for interpolation methods) – is defined exactly like the approximating function of conventional collocation methods. According to eqn. (7.14) it reads as follows:

$$\boldsymbol{y}_k(t) \overset{\text{def}}{=} \sum_{l=0}^{q} \frac{1}{l!}\,(t-t_k)^l\,\boldsymbol{y}_{k0}^{(l)} \; .$$

The most efficient algorithms result when assuming the collocation epochs to be equidistantly spaced within each subinterval. Because the collocation epochs t_{k_j} coincide with subinterval boundaries in the case of multistep methods (see Figure 7.5), *an equidistant spacing of the collocation epochs implies an equidistant spacing of all subinterval boundaries* t_k. Therefore multistep methods with an equidistant spacing of collocation epochs are methods of constant stepsize h, where

$$h = t_{k+1} - t_k \overset{\text{def}}{=} \text{const.} \ . \tag{7.97}$$

From eqns. (7.17) and (7.18) we conclude that the collocation epochs are defined as follows:

$$t_{k_j} \overset{\text{def}}{=} t_k - (j - m)\, h = t_{k-j+m} \ , \quad j = 1, 2, \ldots, q + 1 - n \ , \tag{7.98}$$

where the method-dependent integer number m is

$$m = \begin{cases} 1 & \text{for extrapolation methods} \\ 2 & \text{for interpolation methods} \end{cases} \ . \tag{7.99}$$

The system (7.15) of condition equations for multistep methods looks as follows:

$$\begin{aligned}
\sum_{l=n}^{q} \frac{(t_{k_j} - t_k)^{l-n}}{(l-n)!}\, \boldsymbol{y}_{k0}^{(l)} \qquad &= \boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k(t_{k_j}), \dot{\boldsymbol{y}}_k(t_{k_j}), \ldots, \boldsymbol{y}_k^{(n-1)}(t_{k_j})\Big) \\
\sum_{l=n}^{q} (j - m)^{l-n} \, \frac{(-h)^{l-n}}{(l-n)!}\, \boldsymbol{y}_{k0}^{(l)} &= \boldsymbol{f}\Big(t_{k_j}, \boldsymbol{y}_k(t_{k_j}), \dot{\boldsymbol{y}}_k(t_{k_j}), \ldots, \boldsymbol{y}_k^{(n-1)}(t_{k_j})\Big) \\
\sum_{l=n}^{q} (-1)^{l-n} \, (j - m)^{l-n} \, \boldsymbol{d}_{k_l} \quad &= \boldsymbol{f}\Big(t_{k-j+m}, \boldsymbol{y}_k(t_{k-j+m}), \ldots, \boldsymbol{y}_k^{(n-1)}(t_{k-j+m})\Big) \\
& \qquad\qquad j = 1, 2, \ldots, q + 1 - n \ ,
\end{aligned} \tag{7.100}$$

where the coefficients $\boldsymbol{d}_{k_l}$ obviously are defined by

$$\boldsymbol{d}_{k_l} \overset{\text{def}}{=} \frac{h^{l-n}}{(l-n)!}\, \boldsymbol{y}_{k0}^{(l)} \ . \tag{7.101}$$

The similarities between the above system of condition equations with that of the conventional collocation methods (eqns. (7.60)) are striking. Therefore, the solution of eqns. (7.100) (using the coefficients $\boldsymbol{d}_{k_l}$ as auxiliary unknowns) may be done in strict analogy to the case of the conventional collocation methods. It makes in particular sense to base the algorithm on the *backwards differences* of the accelerations $\boldsymbol{f}(t_{k-j+m})$, $j = 1, 2, \ldots, q+1-n$ (as opposed to the *forward differences* considered in section 7.5.3). This implies, however, that the backward differences, as illustrated by Figure 7.17 in the case of the extrapolation methods, are used. These backward differences of the function values $\boldsymbol{f}(t_{k-j+m})$, $j = 1, 2, \ldots, q + 1 - n$, are defined by the equations

$$\begin{aligned}
\nabla_{k-j+m}^{[0]} &\overset{\text{def}}{=} \boldsymbol{f}(t_{k-j+m}) \\
\nabla_{k-j+m}^{[l]} &\overset{\text{def}}{=} \nabla_{k-j+m}^{[l-1]} - \nabla_{k-j+m-1}^{[l-1]} \ .
\end{aligned} \tag{7.102}$$

$$\begin{array}{lccccccc}
\boldsymbol{f}(t_{k_7}) = \boldsymbol{f}(t_{k-6}) \overset{\text{def}}{=} \nabla^{[0]}_{k-6} & & & & & & & \\
& \nabla^{[1]}_{k-5} & & & & & & \\
\boldsymbol{f}(t_{k_6}) = \boldsymbol{f}(t_{k-5}) \overset{\text{def}}{=} \nabla^{[0]}_{k-5} & & \nabla^{[2]}_{k-4} & & & & & \\
& \nabla^{[1]}_{k-4} & & \nabla^{[3]}_{k-3} & & & & \\
\boldsymbol{f}(t_{k_5}) = \boldsymbol{f}(t_{k-4}) \overset{\text{def}}{=} \nabla^{[0]}_{k-4} & & \nabla^{[2]}_{k-3} & & \nabla^{[4]}_{k-2} & & & \\
& \nabla^{[1]}_{k-3} & & \nabla^{[3]}_{k-2} & & \nabla^{[5]}_{k-1} & & \\
\boldsymbol{f}(t_{k_4}) = \boldsymbol{f}(t_{k-3}) \overset{\text{def}}{=} \nabla^{[0]}_{k-3} & & \nabla^{[2]}_{k-2} & & \nabla^{[4]}_{k-1} & & \nabla^{[6]}_{k} & \\
& \nabla^{[1]}_{k-2} & & \nabla^{[3]}_{k-1} & & \nabla^{[5]}_{k} & & \\
\boldsymbol{f}(t_{k_3}) = \boldsymbol{f}(t_{k-2}) \overset{\text{def}}{=} \nabla^{[0]}_{k-2} & & \nabla^{[2]}_{k-1} & & \nabla^{[4]}_{k} & & & \\
& \nabla^{[1]}_{k-1} & & \nabla^{[3]}_{k} & & & & \\
\boldsymbol{f}(t_{k_2}) = \boldsymbol{f}(t_{k-1}) \overset{\text{def}}{=} \nabla^{[0]}_{k-1} & & \nabla^{[2]}_{k} & & & & & \\
& \nabla^{[1]}_{k} & & & & & & \\
\boldsymbol{f}(t_{k_1}) = \boldsymbol{f}(t_{k}) \quad \overset{\text{def}}{=} \nabla^{[0]}_{k} & & & & & & &
\end{array}$$

Fig. 7.17. Visualization of backward differences up to order 6 for extrapolation methods

After analogous transformations as in the case of collocation, the coefficients $\boldsymbol{d}_{k_l}$ for a *pure extrapolation algorithm* may be written in the following convenient matrix form:

$$\mathbf{D}_k^{I+1} = \tilde{\mathbf{N}}^{-1}\, \tilde{\mathbf{F}}_k^{I} \ , \tag{7.103}$$

where $\mathbf{D}_k^{I+1}$ is defined in analogy to the first of eqns. (7.64), matrix $\tilde{\mathbf{N}}^{-1}$ by eqn. (7.104)

$$\tilde{\mathbf{N}}^{-1} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} \\
0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{11}{24} & \frac{5}{12} & \frac{137}{360} & \frac{7}{20} & \frac{363}{1120} & \frac{761}{2520} & \frac{7129}{25200} \\
0 & 0 & 0 & \frac{1}{6} & \frac{1}{4} & \frac{7}{24} & \frac{5}{16} & \frac{29}{90} & \frac{469}{1440} & \frac{29531}{90720} & \frac{1303}{4032} \\
0 & 0 & 0 & 0 & \frac{1}{24} & \frac{1}{12} & \frac{17}{144} & \frac{7}{48} & \frac{967}{5760} & \frac{89}{480} & \frac{4523}{22680} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{120} & \frac{1}{48} & \frac{5}{144} & \frac{7}{144} & \frac{1069}{17280} & \frac{19}{256} \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{720} & \frac{1}{240} & \frac{23}{2880} & \frac{1}{80} & \frac{3013}{172800} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5040} & \frac{1}{1440} & \frac{13}{8640} & \frac{1}{384} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{40320} & \frac{1}{10080} & \frac{29}{120960} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{362880} & \frac{1}{80640} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3628800}
\end{pmatrix} \tag{7.104}$$

and

$$\tilde{\mathbf{F}}_k^I = \begin{pmatrix} \left[\boldsymbol{f}^I(t_k)\right]^T \\ \left[\nabla_k^{[1]}\right]^T \\ \left[\nabla_k^{[2]}\right]^T \\ \cdots \\ \cdots \\ \cdots \\ \left[\nabla_k^{[q-n]}\right]^T \end{pmatrix} . \tag{7.105}$$

Using the defining equation (7.100) one may easily verify that matrix $\tilde{\mathbf{N}}^{-1}$ has to be replaced by eqn. (7.106) for an *interpolation algorithm*

$$\tilde{\mathbf{N}}_{\text{int}}^{-1} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{-1}{2} & \frac{-1}{6} & \frac{-1}{12} & \frac{-1}{20} & \frac{-1}{30} & \frac{-1}{42} & \frac{-1}{56} & \frac{-1}{72} & \frac{-1}{90} \\ 0 & 0 & \frac{1}{2} & \frac{0}{1} & \frac{-1}{24} & \frac{-1}{24} & \frac{-13}{360} & \frac{-11}{360} & \frac{-29}{1120} & \frac{-223}{10080} & \frac{-481}{25200} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{12} & \frac{1}{24} & \frac{1}{48} & \frac{7}{720} & \frac{1}{288} & \frac{-1}{5670} & \frac{-61}{25920} \\ 0 & 0 & 0 & 0 & \frac{1}{24} & \frac{1}{24} & \frac{5}{144} & \frac{1}{36} & \frac{127}{5760} & \frac{101}{5760} & \frac{1271}{90720} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{120} & \frac{1}{80} & \frac{1}{72} & \frac{1}{72} & \frac{229}{17280} & \frac{427}{34560} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{720} & \frac{1}{360} & \frac{11}{2880} & \frac{13}{2880} & \frac{853}{172800} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5040} & \frac{1}{2016} & \frac{7}{8640} & \frac{19}{17280} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{40320} & \frac{1}{13440} & \frac{17}{120960} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{362880} & \frac{1}{103680} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3628800} \end{pmatrix} \tag{7.106}$$

and that the backward differences $\tilde{\mathbf{F}}_k^I$ referring to epoch t_k have to be replaced by those referring to epoch t_{k+1}:

$$\tilde{\mathbf{G}}_k^I = \begin{pmatrix} \left[\boldsymbol{f}^I(t_{k+1})\right]^T \\ \left[\nabla_{k+1}^{[1]}\right]^T \\ \left[\nabla_{k+1}^{[2]}\right]^T \\ \cdots \\ \cdots \\ \cdots \\ \left[\nabla_{k+1}^{[q-n]}\right]^T \end{pmatrix} . \tag{7.107}$$

The coefficients in the case of interpolation read as

$$\mathbf{D}_k^{I+1} = \tilde{\mathbf{N}}_{\text{int}}^{-1}\,\tilde{\mathbf{G}}_k^I \ . \tag{7.108}$$

A multistep procedure (extrapolation and interpolation) is included in the program package accompanying this book. The programs NUMINT and

PLASYS are capable of performing the integration either with a conventional collocation method as discussed in section 7.5 or with a multistep procedure. If only orbits with small eccentricities (e.g., the planetary system without Mercury and Pluto) are integrated, the multistep procedure is an excellent, probably even the best possible, choice. It is, however, the program user's task to define a suitable fixed stepsize (see section 7.7.4 for guidelines).

The diagonal of matrix $\tilde{\mathbf{N}}^{-1}$ obviously contains the terms $1/l!\,,\ \ l = 1, 2, \ldots, q-n\,.$ As the lower diagonal part of the matrix only contains zero elements and as the terms $\boldsymbol{d}_{k_l}$ are defined by eqn. (7.101), we may conclude that

$$\boldsymbol{y}^{(q)}_{I+1\,k0} = h^{-(q-n)}\ \nabla_k^{[q-n]} + \boldsymbol{O}(h_k)\ . \tag{7.109}$$

The above equation says that the $(q-n)$-th backward difference of the forces may be interpreted (in a modest approximation and apart from a scaling factor) as the q-th derivative of the solution vector $\boldsymbol{y}_k(t)$ at t_k .

Equation (7.103) tells that the coefficients $\boldsymbol{d}_{k_l}\,,\ l = n, n+1, \ldots, q$, are linear combinations of the backward differences $\nabla_k^{[l]}\,,\ \ l = 0, 1, \ldots, q-n\,.$ Consequently (see eqn. (7.100)), the solution vector $\boldsymbol{y}_k(t)$ may be represented as a linear combination of the same differences. The task is achieved by introducing the interval-independent time argument

$$\tau = \frac{1}{h}\,(t - t_k)\ . \tag{7.110}$$

Proceeding in an analogous way as in the case of the conventional collocation method we obtain:

$$\begin{aligned}\boldsymbol{y}_k(t) &= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t-t_k)^l\,\boldsymbol{y}^{(l)}_{k0}\\ &\quad + \sum_{j=n}^{q}\left\{\sum_{l=n}^{j}\frac{(l-n)!}{l!}\,\tilde{\mathrm{N}}^{-1}_{l-n+1,j-n+1}\,\tau^l\right\} h^n\,\nabla_k^{[j-n]}\ .\end{aligned} \tag{7.111}$$

Observe that the term in brackets $\{\ldots\}$ does not contain any terms specific for the subinterval I_k . For the purpose of error control we do not only need the formula for the solution vector, but also for its first $n-1$ derivatives. Taking into account that

$$\frac{d^i}{dt^i} = \left(\frac{q-n}{h_k}\right)^i \frac{d^i}{d\tau^i}\ , \tag{7.112}$$

we obtain the result:

$$\begin{aligned}\boldsymbol{y}_k^{(i)}(t) &= \sum_{l=i}^{n-1} \frac{1}{(l-i)!}\,(t-t_k)^{l-i}\,\boldsymbol{y}_{k0}^{(l)} \\ &+ \sum_{j=n}^{q} \left\{ \sum_{l=n}^{j} \frac{(l-n)!}{(l-i)!}\,\tilde{\mathrm{N}}_{l-n+1,j-n+1}^{-1}\,\tau^{l-i} \right\} h^{n-i}\ \nabla_k^{[j-n]} \\ &\qquad\qquad i = 0,1,\ldots,n-1\ .\end{aligned} \tag{7.113}$$

The multistep method provided in the program system contains the coefficient matrices up to degree and order $q-n = 14$. The maximum order thus would be $q = 16$ for a differential equation system of order $n = 2$. The algorithm supplied in the program system is able to handle any order $n \geq 1$ of the differential equation system. It may be initialized either by the Euler approximation or by the numerical solution of a previous step.

As already pointed out in section 7.4.2 only the function values $\boldsymbol{f}(t_{k+1})$ are (re)calculated in the integration step referring to t_k as initial epoch for $k > 0$. The function values $\boldsymbol{f}(t_{k+1-j})$, $j = 1,2,\ldots,q$, are taken over without any changes from the previous interval I_{k-1}. As a matter of fact, the backward differences and not the function values themselves are updated, where the following scheme is used:

$$\begin{aligned}\nabla_{k+1}^{[0]} &\overset{\text{def}}{=} \boldsymbol{f}(t_{k+1}) \\ \nabla_{k+1}^{[l+1]} &= \nabla_{k+1}^{[l]} - \nabla_k^{[l]}\ , \quad l = 0,1,\ldots,q-1-n\ .\end{aligned} \tag{7.114}$$

Table 7.3 gives an impression of the performance of both, the classical collocation and the multistep methods, implemented in the program package. The stepsizes $h \overset{\text{def}}{=} t_{k+1} - t_k$, together with the order q of the method, are the independent arguments in Table 7.3. One might at first sight think that the collocation method is much better than the multistep method, because the stepsizes h are much longer in the former case. This is not true, however: The number defining efficiency (at least for complicated differential equation systems) is the number of evaluations of the right-and sides $\boldsymbol{f}(t)$ of the differential equation systems in a given time interval. The number n_f of evaluations per revolution period of the planet is therefore included, as well, in Table 7.3.

How do we have to select the stepsize h of the multistep step method to achieve (roughly) the same performance as the corresponding collocation method? Keeping in mind that the coefficients of the approximating function (7.14) are obtained as solutions of the condition equations (7.15) one would expect that the length of the interval containing all $q+1-n$ collocation epochs should be identical for both, the collocation and the multistep method. Therefore we expect a similar performance for

Table 7.3. Multistep methods and conventional collocation methods

	Multistep				Collocation			
q	h	n_f	δa	δu	h	n_f	δa	δu
	[days]		[AU]	[deg]	[days]		[AU]	[deg]
10	10.0	145	$2.4 \cdot 10^{-8}$	$6.8 \cdot 10^{-4}$	80	307	$1.2 \cdot 10^{-9}$	$3.0 \cdot 10^{-5}$
10	10.0	289	$7.0 \cdot 10^{-10}$	$2.0 \cdot 10^{-5}$	80	451	$5.0 \cdot 10^{-12}$	$1.1 \cdot 10^{-7}$
11	9.0	161	$4.0 \cdot 10^{-10}$	$9.0 \cdot 10^{-6}$	80	343	$4.7 \cdot 10^{-11}$	$1.2 \cdot 10^{-6}$
11	9.0	321	$7.7 \cdot 10^{-12}$	$1.5 \cdot 10^{-7}$	80	506	$7.5 \cdot 10^{-13}$	$7.0 \cdot 10^{-9}$
12	8.0	181	$2.7 \cdot 10^{-11}$	$8.5 \cdot 10^{-7}$	80	379	$1.5 \cdot 10^{-11}$	$4.0 \cdot 10^{-7}$
12	8.0	361	$7.0 \cdot 10^{-13}$	$2.2 \cdot 10^{-8}$	80	560	$7.0 \cdot 10^{-14}$	$1.4 \cdot 10^{-9}$
13	7.0	207	$3.0 \cdot 10^{-13}$	$4.0 \cdot 10^{-9}$	80	415	$4.4 \cdot 10^{-12}$	$1.2 \cdot 10^{-7}$
13	7.0	413	$9.0 \cdot 10^{-14}$	$8.0 \cdot 10^{-10}$	80	614	$8.0 \cdot 10^{-14}$	$2.0 \cdot 10^{-9}$
14	7.0	207	$7.0 \cdot 10^{-14}$	$2.0 \cdot 10^{-9}$	80	451	$8.5 \cdot 10^{-11}$	$2.2 \cdot 10^{-6}$
14	7.0	413	$6.0 \cdot 10^{-14}$	$7.3 \cdot 10^{-9}$	80	668	$6.0 \cdot 10^{-13}$	$1.6 \cdot 10^{-8}$
10	12.5	116	$1.6 \cdot 10^{-7}$	$5.0 \cdot 10^{-3}$	100	246	$2.5 \cdot 10^{-8}$	$8.0 \cdot 10^{-4}$
10	12.5	231	$5.1 \cdot 10^{-9}$	$1.4 \cdot 10^{-4}$	100	361	$5.5 \cdot 10^{-11}$	$1.1 \cdot 10^{-6}$
11	11.0	132	$3.7 \cdot 10^{-9}$	$8.0 \cdot 10^{-5}$	100	274	$2.5 \cdot 10^{-9}$	$7.0 \cdot 10^{-5}$
11	11.0	263	$7.7 \cdot 10^{-11}$	$1.6 \cdot 10^{-6}$	100	404	$3.3 \cdot 10^{-11}$	$7.0 \cdot 10^{-7}$
12	10.0	145	$3.0 \cdot 10^{-10}$	$1.0 \cdot 10^{-5}$	100	303	$4.4 \cdot 10^{-10}$	$1.2 \cdot 10^{-5}$
12	10.0	289	$7.5 \cdot 10^{-12}$	$2.4 \cdot 10^{-7}$	100	448	$7.5 \cdot 10^{-12}$	$2.6 \cdot 10^{-7}$
13	9.0	161	$6.0 \cdot 10^{-12}$	$1.2 \cdot 10^{-7}$	100	332	$8.0 \cdot 10^{-11}$	$1.8 \cdot 10^{-6}$
13	9.0	321	$1.5 \cdot 10^{-13}$	$4.0 \cdot 10^{-9}$	100	491	$7.5 \cdot 10^{-13}$	$2.0 \cdot 10^{-8}$
14	8.0	181	$5.0 \cdot 10^{-13}$	$2.0 \cdot 10^{-8}$	100	361	$9.0 \cdot 10^{-11}$	$2.5 \cdot 10^{-6}$
14	8.0	361	$1.0 \cdot 10^{-13}$	$4.0 \cdot 10^{-9}$	100	535	$5.0 \cdot 10^{-13}$	$1.2 \cdot 10^{-8}$
12	12.0	121	$2.0 \cdot 10^{-9}$	$6.8 \cdot 10^{-5}$	120	253	$2.0 \cdot 10^{-9}$	$5.0 \cdot 10^{-5}$
12	12.0	241	$5.0 \cdot 10^{-11}$	$1.5 \cdot 10^{-6}$	120	373	$8.4 \cdot 10^{-11}$	$2.2 \cdot 10^{-6}$
13	11.0	132	$9.0 \cdot 10^{-11}$	$2.2 \cdot 10^{-6}$	120	277	$2.8 \cdot 10^{-9}$	$7.0 \cdot 10^{-5}$
13	11.0	263	$1.6 \cdot 10^{-12}$	$3.7 \cdot 10^{-8}$	120	409	$7.5 \cdot 10^{-12}$	$2.0 \cdot 10^{-7}$
14	10.0	–	–	–	120	301	$5.0 \cdot 10^{-11}$	$8.0 \cdot 10^{-7}$
14	10.0	289	$2.2 \cdot 10^{-13}$	$8.2 \cdot 10^{-9}$	120	445	$1.8 \cdot 10^{-12}$	$5.0 \cdot 10^{-8}$

$$h_{\text{multistep}} \approx \frac{1}{q-n}\, h_{\text{collocation}} \ . \tag{7.115}$$

Tests with three different collocation stepsizes, namely $h = 80,\ 100,\ 120$ days, were made. According to the above rule this should roughly correspond to steps of order dependent size $h_{\text{mulitstep}} = h_{\text{collocation}} \,/(q-2)$ in the case of multistep methods. In order to avoid numerical problems, the resulting steps $h_{\text{multistep}}$ were rounded to an integer number of days. The integration order q was varied within the limits $10 \leq q \leq 14$. These integration orders are reasonable for orbital dynamics (planetary system and satellite geodesy) and a floating point environment with 14 hexadecimal digits.

For a particular stepsize h one expects that the actual errors are governed by the *approximation error* for the lower orders q, by *rounding errors* for the higher orders.

For both methods the error in the semi-major axis a (in AU) and in the argument of latitude (in °) are tabulated (the error in the position would result after multiplication with the factor $\approx \frac{\pi}{180}\, a$).

For both methods, the number of iterative improvements of the solution per subinterval I_k was varied: The first line for each order q corresponds to a pure extrapolation method in the case of the multistep method and to one iteration step in the case of conventional collocation methods. The second line corresponds to a multistep method with one interpolation step and to a collocation method with two iteration steps.

Let us now discuss the results summarized in Table 7.3. First of all, we should point out that all results in Table 7.3 are of a good quality. An inspection of the numbers n_f and the corresponding accuracies indicates that the multistep procedure is about a factor of $1.5 - 3$ more efficient than the collocation method for these comparatively high integration orders. Multistep methods are clearly preferable if the numerical solution is dominated by rounding errors. The integration failed in one case, $q = 14$ for $h = 10$ in the case of the multistep method and pure extrapolation, indicating that in the computing environment given (double precision floating point) one should not try to select very high orders and very long stepsizes.

One big advantage of the conventional collocation methods over the multistep methods resides in the fact that they may be easily modified to allow for automatic stepsize control. We have shown in section 7.5.5 that stepsize control may be achieved using the highest term $\boldsymbol{y}_{k0}^{(q)}$ of the approximating function in the case of conventional collocation methods. It would not be difficult, in principle, to transform a multistep method into a procedure allowing for stepsize control, as well. But a change of stepsize would require a re-evaluation of all function values $\boldsymbol{f}(t_{k_j})$, $j = 1, 2, \ldots, q + 1 - n$, and not only of the last one. Such a procedure, although it would work perfectly, would be rather inefficient. This is why efficient stepsize control in the case of multistep methods has to be performed by controlling the spacing h_k between the subinterval boundaries t_k. There are no problems of principle involved in such a procedure: Our discussion of collocation methods has shown that equal spacing between the epochs t_k is not a requirement, but that it leads to very efficient algorithms. By dropping the requirement of an equal spacing, a good part of the simplicity and elegance of multistep methods is lost. Nevertheless, very powerful methods were developed for first order systems. For a profound discussion we refer, e.g., to [108]. In view of the simplicity and efficiency of the conventional collocation algorithm we do not use multistep algorithms with stepsize control in this book.

Table 7.3 indicates, on the other hand, that high-order multistep methods ($10 \leq q \leq 14$) should be given the preference over the conventional collocation methods, if only orbits of small eccentricities are considered.

Many "different" multistep procedures are distinguished in the literature. The method developed here is equivalent to the so-called Adams-Bashford method (extrapolation) and to the Adams-Bashford-Moulton method (interpolation), when applied to first order differential equation. When applied to second-order differential equation systems our multistep method is equivalent, from the algebraical point of view, to the so-called Stormer method (extrapolation) and the Cowell method (interpolation).

In the original Stormer and Cowell algorithms the position and velocity vector (in the case of a general differential equation of order n: the derivatives $\boldsymbol{y}^{(i)}$, $i = 1, 2, \ldots, n{-}1$) are replaced by the differences $\nabla \boldsymbol{y}^{[i]}$, $i = 1, 2, \ldots n{-}1$, of the solution vector w.r.t. the current initial epoch (the differences are formed in analogy of those of the vector $\boldsymbol{f}(t_{k_j})$). The advantage of this formulation over the one we use resides in the hierarchy of the differences: one may assume that the differences of order i are (at least) one order of magnitude smaller in absolute value than those of order $i - 1$. This implies that there is also a hierarchy in the absolute value of rounding errors when computing these differences. Despite these apparent advantages we did not use the original Stormer and Cowell formulations, in order to keep the algorithms simple and general (applicable to equations of all orders $n > 0$).

7.6 Linear Differential Equation Systems and Numerical Quadrature

7.6.1 Introductory Remarks

Linear differential equation systems form a special class of ordinary differential equation systems, which may be solved by any method introduced so far (capable of solving general, non-linear and linear, systems of equations). Strictly speaking, it is therefore not necessary to discuss special methods for linear differential equation systems. The readers not interested in such *subtleties* may skip the entire section 7.6 and continue reading section 7.7.

Taylor series methods and collocation methods are capable of exploiting the linearity of systems or the fact that the right-hand sides of the system are merely known functions of time t. Neither the Runge-Kutta nor extrapolation methods are candidates to solve such problems.

Exactly as in the general case, it may be necessary to divide the original integration interval I into subintervals I_k , $k = 0, 1, 2, \ldots$. As the transition from one subinterval to the next is performed in the same way for linear and non-linear system, we need not address this issue subsequently. It is therefore perfectly allowed to skip the subinterval index k in this section.

The problem to be studied in this section may be written either as initial or as boundary value problem. Let us denote it by:

$$\boldsymbol{y}^{(n)} = \boldsymbol{f}\big(t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ldots, \boldsymbol{y}^{(n-1)}\big) = \sum_{i=0}^{n-1} \mathbf{A}_i(t)\, \boldsymbol{y}^{(i)} \;+\; \boldsymbol{b}(t) \;, \tag{7.116}$$

Initial value problem:

$$\boldsymbol{y}^{(i)}(t_0) \overset{\text{def}}{=} \boldsymbol{y}_0^{(i)}, \qquad i = 0, 1, \ldots, n-1 \;. \tag{7.117}$$

Boundary value problem:

$$\boldsymbol{y}^{(k_i)}(t_i') \overset{\text{def}}{=} \boldsymbol{y}_i^{(k_i)}, \qquad i = 1, 2, \ldots, n \,, \quad k_i \in \{0, 1, \ldots, n-1\} \;, \tag{7.118}$$

where the above boundary value problem is a slight generalization of the problem (7.6), which allowed only for zero-order derivatives, i.e., for $k_i = 0$, $i = 1, 2, \ldots, n$. Observe that not all boundary value problems of type (7.116, 7.118) may be solved. It is, e.g., a requirement that at least for one index i we have $k_i = 0$. The coefficient matrices $\mathbf{A}_i(t)$ are square matrices of dimension d (d is the dimension of the system), the inhomogeneous part $\boldsymbol{b}(t)$ is a column matrix of dimension d.

7.6.2 Taylor Series Solution

The numerical solution of the initial value problem (7.116, 7.117) (not of the more general case) is sought in the form

$$\begin{aligned} \boldsymbol{y}(t) &= \sum_{l=0}^{q} \frac{1}{l!}\,(t-t_0)^l\, \boldsymbol{y}_0^{(l)} \\ &= \sum_{l=0}^{n-1} \frac{1}{l!}\,(t-t_0)^l\, \boldsymbol{y}_0^{(l)} \;+\; \sum_{l=n}^{q} \frac{1}{l!}\,(t-t_0)^l\, \boldsymbol{f}^{(l-n)}(t) \;. \end{aligned} \tag{7.119}$$

The first n coefficients of the series are defined by the initial condition. The derivatives of n-th and higher order may be calculated as follows:

$$
\begin{aligned}
\boldsymbol{y}^{(n)}(t_0) \quad &= \sum_{i=0}^{n-1} \mathbf{A}_i(t_0)\,\boldsymbol{y}^{(i)}(t_0) \,+\, \boldsymbol{b}(t_0) \\
\boldsymbol{y}^{(n+1)}(t_0) &= \sum_{i=0}^{n-1} \mathbf{A}_i(t_0)\,\boldsymbol{y}^{(i+1)}(t_0) \,+\, \sum_{i=0}^{n-1} \dot{\boldsymbol{A}}_i(t_0)\,\boldsymbol{y}^{(i)}(t_0) \,+\, \dot{\boldsymbol{b}}(t_0) \\
\boldsymbol{y}^{(n+2)}(t_0) &= \sum_{i=0}^{n-1} \mathbf{A}_i(t_0)\,\boldsymbol{y}^{(i+2)}(t_0) \,+\, 2\sum_{i=0}^{n-1} \dot{\boldsymbol{A}}_i(t_0)\,\boldsymbol{y}^{(i+1)}(t_0) \\
&\quad + \sum_{i=0}^{n-1} \ddot{\boldsymbol{A}}_i(t_0)\,\boldsymbol{y}^{(i)}(t_0) \,+\, \ddot{\boldsymbol{b}}(t_0) \\
\ldots \quad &.
\end{aligned}
\tag{7.120}
$$

Equations (7.120) together with the initial conditions in eqns. (7.117) define an algorithm to compute the higher-order derivatives. The information needed to compute derivative number i is available, if all derivatives of lower order are available.

This statement is only true, however, if the matrices $\mathbf{A}_i^{(l)}(t_0),\ l = 0, 1, \ldots, q$, and $\boldsymbol{b}(t_0)$ are easily available. In general this will not be the case. As opposed to the non-linear case, there is, however, an easy way to solve this problem with an accuracy of order $\boldsymbol{O}((t - t_0)^q)$, simply by replacing the matrix elements by their interpolating polynomials of degree $q - n$ defined by the function values $\mathbf{A}_i(t_{0_j}),\ j = 1, 2, \ldots, q + 1 - n$. The interpolation epochs t_{0_j} (in principle) may be selected arbitrarily, provided all epochs are different.

This procedure is sufficient to make the algorithm defined by eqns. (7.120) one of order q for $\boldsymbol{y}(t)$ in $t - t_0$ (if the series is terminated after the terms of order q). The resulting algorithm is rather efficient, if not only one, but many initial value problems referring to one and the same homogeneous part of the differential equation system (7.116), differing "only" by their initial values and/or the non-homogeneous parts $\boldsymbol{b}$, have to be solved. This is, e.g., the case, if many variational equations referring to the same primary equations have to be integrated.

The above formulae are drastically reduced when applied to the numerical solution of a definite integral. In this case the higher-order derivatives simply are computed as

$$
\boldsymbol{y}^{(i)}(t_0) = \boldsymbol{f}^{(i-n)}(t_0)\,, \quad i = n, n+1, \ldots, q\ . \tag{7.121}
$$

It is now even possible to deal with each component of the solution vector separately, i.e., the problem is split up into the solution of d separate integrals. If the (mathematically) correct formulae for the derivatives $\boldsymbol{f}(t_0)^{(i)},\ i = 1, 2, \ldots, q$, are not available (and this is the general case), these derivatives must be replaced by the derivatives of an interpolating polynomial. The above algorithm might be further refined. For more information we refer to [15].

7.6.3 Collocation for Linear Systems: Basics

Collocation methods to solve the linear initial value problem (7.116, 7.117) make explicit use of the linearity of the system (7.15) of condition equations when replacing the iterative process (7.51) by a direct solution of the linear system (7.15) in one step.

This system shall be given in explicit form. In a first step the order of the double sum on the right-hand side of the differential equation system is reversed (for one of the equations of the system (7.15) with time argument t):

$$\begin{aligned}\sum_{l=n}^{q} \frac{(t-t_0)^{l-n}}{(l-n)!}\, \boldsymbol{y}_0^{(l)} &= \sum_{i=0}^{n-1} \mathbf{A}_i(t) \sum_{l=i}^{q} \frac{(t-t_0)^{l-i}}{(l-i)!}\, \boldsymbol{y}_0^{(l)} + \boldsymbol{b}(t) \\ &= \sum_{l=0}^{q} \sum_{i=0}^{\tilde{n}} \frac{(t-t_0)^{l-i}}{(l-i)!}\, \mathbf{A}_i(t)\, \boldsymbol{y}_0^{(l)} + \boldsymbol{b}(t) \\ &\quad \tilde{n} = l \text{ for } l < n,\ \tilde{n} = n-1 \text{ for } l \geq n \ .\end{aligned} \tag{7.122}$$

Using the notations

$$\begin{aligned}\tilde{\mathbf{A}}_n &= \mathbf{E} \\ \tilde{\mathbf{A}}_i &= -\,(t-t_0)^{n-i}\, \mathbf{A}_i \ , \quad i = 0, 1, \ldots, n-1 \ ,\end{aligned} \tag{7.123}$$

where $\mathbf{E}$ is the unit matrix of dimension d, eqns. (7.122) may be written as:

$$\sum_{l=0}^{q} (t-t_0)^{l-n} \left[\sum_{i=0}^{\tilde{n}} \frac{\tilde{\mathbf{A}}_i(t)}{(l-i)!} \right] \boldsymbol{y}_0^{(l)} = \boldsymbol{b}(t) \ , \quad \tilde{n} = l \text{ for } l \leq n, \quad \tilde{n} = n \text{ else.} \tag{7.124}$$

The coefficients are determined by the request that either the initial or the boundary conditions are met and that eqns. (7.124) hold at $q+1-n$ different epochs t_{0_j} , $j = 1, 2, \ldots, q+1-n$. The condition equations may thus be written as:

For the initial value problem:

$\boldsymbol{y}_0^{(i)} = \boldsymbol{y}_0^{(i)}(t_0)$, $\quad i = 0, 1, \ldots, n-1$.

For the boundary value problem:

$\displaystyle\sum_{l=k_i}^{q} \frac{(t_i' - t_0)^{l-k_i}}{(l-k_i)!}\, \boldsymbol{y}_0^{(i)} = \boldsymbol{y}_i^{(k_i)}$, $\quad i = 1, 2, \ldots, n \quad k_i \in \{0, 1, \ldots, n-1\}$.

Collocation conditions:

$$\sum_{l=0}^{q}(t_{0_j}-t_0)^{l-n}\left[\sum_{i=0}^{\tilde{n}}\frac{\tilde{\mathbf{A}}_i(t_{0_j})}{(l-i)!}\right]\boldsymbol{y}_0^{(l)}=\boldsymbol{b}(t_{0_j})\,,\quad \tilde{n}=l \text{ for } l\le n,\quad \tilde{n}=n \text{ else}$$
$$j=1,2,\ldots,q+1-n\,, \tag{7.125}$$

where the epochs t'_i, $i=1,2,\ldots,n-1$, are the boundary epochs, and k_i is the derivative specified at epoch t'_i. t_0 is the initial epoch. The coefficients $\boldsymbol{y}_0^{(l)}$, $l=0,1,\ldots,q$, are the solutions of the above system of linear algebraic equations.

The algorithm defined by eqns. (7.125) deserves a few comments:

- Algorithms of type (7.125) are the classical collocation methods encountered in the literature. Subroutine libraries like the NAg-Library [81] contain routines based on eqns. of this type.
- As opposed to the application of collocation algorithms to non-linear systems,
  - we have to solve one linear system of equations of dimension $d\ (q+1)$ in one step *and not* d separate systems of dimension $q+1$ iteratively,
  - the differential equation systems hold *exactly* (apart from rounding errors) at the $q+1-n$ epochs t_{0_j}, $j=1,2,\ldots,q+1-n$, and not only up to the order q in $(t-t_0)$,
  - the resulting matrix of coefficients for the determination of the unknowns $\boldsymbol{y}_0^{(i)}$, $i=0,1,\ldots,q$, is problem-dependent, i.e., there is no way to separate the inversion of the coefficients from the actual problem (as it could be done in the general case) and to compute the inverses a priori.
- Initial- or boundary-value problems referring to the same homogeneous part of the linear system (7.116), i.e., to the same matrices $\mathbf{A}_i$, may be dealt with in a more efficient way, because the matrix inversion is required only once for all particular solutions.
- For initial value problems the first n coefficients are obtained directly from the initial conditions. The remaining coefficients $\boldsymbol{y}_0^{(i)}$, $i=n,n+1,\ldots,q$, solve the system of the $d\ (q+1-n)$ last equations in the algorithm (7.125).
- The solution of the initial value problem associated with the homogeneous system (i.e., for which $\boldsymbol{b}(t)=\mathbf{0}$) may be represented as a linear combination of the elements of the initial state vector $\boldsymbol{y}_0^{(i)}$, $i=0,1,\ldots,n-1$.
- It is particularly interesting to study the impact of the distribution of the epochs t_{0_j}, $j=1,2,\ldots,q+1-n$, within the integration interval on the quality of the solution vector. This question is intimately related to the structure of the error function and will be studied in the following paragraph.

7.6.4 Collocation: Structure of the Local Error Function

We confine our studies to the initial value problem in this section. So far, there was no need in this Chapter to make a clear distinction between the *true* solution of the initial value problem at t_0 and its *numerical approximation.* This distinction is, however, vital for the following discussion. Therefore, let

$\boldsymbol{y}(t)$ designate the true solution of the initial value problem (7.116, 7.117),

$\boldsymbol{z}(t)$ the numerical solution of the same problem, using a collocation method of order q, and

$\boldsymbol{\varepsilon}(t) \overset{\text{def}}{=} \boldsymbol{z}(t) - \boldsymbol{y}(t)$ the error function (where rounding errors are *not* considered).

Observe that an approximation for the error function was already used in section 7.5.5, eqn. (7.85) for the purpose of automatic stepsize control. This approximation cannot be used here, where we are interested in the true structure of the error function $\boldsymbol{\varepsilon}(t)$.

The error function solves a linear, non-homogeneous system of differential equations, where the homogeneous part is identical with that of the differential equation system in problem (7.116). In order to prove this statement, we make use of the fact that the numerical solution and (of course) the true solution exactly solve the differential equation system at the collocation epochs:

$$\begin{array}{llll}
\boldsymbol{z}^{(n)}(t_{0_j}) = & \sum\limits_{i=0}^{n-1} \mathbf{A}_i(t_{0_j})\,\boldsymbol{z}^{(i)}(t_{0_j}) \;+\; \boldsymbol{b}(t_{0_j}) \,, & & j = 1, 2, \ldots, q+1-n \\
\boldsymbol{y}^{(n)}(t_{0_j}) = & \sum\limits_{i=0}^{n-1} \mathbf{A}_i(t_{0_j})\,\boldsymbol{y}^{(i)}(t_{0_j}) \;+\; \boldsymbol{b}(t_{0_j}) \,, & & j = 1, 2, \ldots, q+1-n \\
\hline
\boldsymbol{\varepsilon}^{(n)}(t_{0_j}) = & \sum\limits_{i=0}^{n-1} \mathbf{A}_i(t_{0_j})\,\boldsymbol{\varepsilon}^{(i)}(t_{0_j}) \,, & & j = 1, 2, \ldots, q+1-n \,.
\end{array} \tag{7.126}$$

The third of the above equations implies that $\boldsymbol{\varepsilon}^{(n)}(t)$ may be written as:

$$\boldsymbol{\varepsilon}^{(n)}(t) = \sum_{i=0}^{n-1} \mathbf{A}_i(t)\,\boldsymbol{\varepsilon}^{(i)}(t) \;+\; \prod_{j=1}^{q+1-n} (t - t_{0_j})\,\boldsymbol{g}(t) \,, \tag{7.127}$$

where $\boldsymbol{g}(t)$ is an analytical function (which may be represented by a Taylor series with origin t_0) – at least if the solution of the original initial value problem is analytical, as well. As the true solution and the numerical approximation obey the same initial conditions, the error function may be represented as the solution of the following initial value problem:

$$
\begin{aligned}
\boldsymbol{\varepsilon}^{(n)}(t) \; &= \sum_{i=0}^{n-1} \mathbf{A}_i(t)\,\boldsymbol{\varepsilon}^{(i)}(t) + \prod_{j=1}^{q+1-n} (t - t_{0_j})\,\boldsymbol{g}(t) \\
\boldsymbol{\varepsilon}^{(i)}(t_0) &= \mathbf{0} \;, \quad i = 0, 1, \ldots, n-1 \;.
\end{aligned}
\tag{7.128}
$$

The structure of the above differential equation is the same as that of a variational equation (5.6) associated with one of the dynamical parameters, which was encountered in Chapter 5. We may therefore write the solution of the initial value problem (7.128) as follows (compare eqns. (5.14) and (5.16)):

$$
\boldsymbol{\varepsilon}^{(i)}(t) = \mathbf{Z}^{(i)}(t)\,\boldsymbol{\alpha}(t) \;, \quad i = 0, 1, \ldots, n-1 \;, \tag{7.129}
$$

where $\mathbf{Z}(t)$ is the rectangular matrix with nd columns and d rows, in which column l contains the elements of the solution $\boldsymbol{\varepsilon}_l(t)$ with index l of the complete system of solutions of the homogeneous system associated with the linear system (7.128). The coefficient matrix $\boldsymbol{\alpha}(t)$ may be expressed by an integral (compare eqn. (5.21)):

$$
\boldsymbol{\alpha}(t) = \int_{t_0}^{t} \tilde{\mathbf{Z}}^{-1}(t')\;\mathbf{F}(t')\;dt' \;, \tag{7.130}
$$

where in our case $\mathbf{F}(t)$ may be written as

$$
\mathbf{F}(t) = \prod_{j=1}^{q+1-n} (t - t_{0_j})\,\boldsymbol{G}(t) \;, \tag{7.131}
$$

where $\boldsymbol{G}(t)$ is a column array with nd elemenent, defined as (compare eqn. (5.20)):

$$
\boldsymbol{G} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \cdots \\ \cdots \\ \mathbf{0} \\ \boldsymbol{g}(t) \end{pmatrix} \;. \tag{7.132}
$$

In order to evaluate the definite integral (7.130) at $t = t_0 + h$ we perform a transformation of the independent argument t, as illustrated by Figure 7.18. Algebraically, the transformation reads as:

$$
\begin{aligned}
\tau \quad &= -1 + \frac{2}{h}\,(t - t_0) \\
t - t_0 &= \frac{h}{2}\,(\tau + 1) \\
dt \quad &= \frac{h}{2}\,d\tau \;.
\end{aligned}
\tag{7.133}
$$

The integral (7.130) thus may be written as:

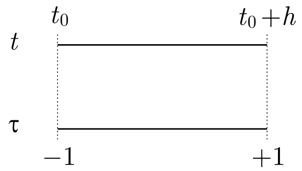


Fig. 7.18. Transformation of the independent argument t

$$\boldsymbol{\alpha}(t_0+h) = \left(\frac{h}{2}\right)^{q+2-n} \int\limits_{-1}^{+1} \prod_{j=1}^{q+1-n} (\tau-\tau_j)\, \tilde{\mathbf{Z}}^{-1}(t'(\tau'))\, \boldsymbol{G}\big(t'(\tau')\big)\, d\tau' \ . \qquad (7.134)$$

This result formally proves that, in general and independently of the selection of the collocation epochs τ_j in the interval $\tilde{I} \overset{\text{def}}{=} [-1,+1]$, the local approximation error is bound by h^{q+2-n} for an integration order q.

The result (7.134) indicates that even a much better approximation may be achieved at $t = t_0 + h$, *provided the collocation epochs τ_j are defined as the roots of the Legendre polynomial $P_{q+1-n}(\tau)$ of degree $q+1-n$.* This can be formally proved by introducing an auxiliary function $\boldsymbol{X}(t(\tau))$ and by representing it as an infinite series of the Legendre polynomials $P_i(\tau)\,,\ i = 0,1,2,\ldots$:

$$\boldsymbol{X}\big(t(\tau)\big) \overset{\text{def}}{=} \tilde{\mathbf{Z}}^{-1}\big(t(\tau)\big)\ \boldsymbol{G}\big(t(\tau)\big) = \sum_{i=0}^{\infty} \left(\frac{h}{2}\right)^{i} \boldsymbol{X}_i\, P_i(\tau)\ . \qquad (7.135)$$

Assuming that the collocation epochs τ_j are the roots of the Legendre polynomial of degree $q+1-n$ we may write this polynomial as

$$\prod_{j=1}^{q+1-n} (\tau-\tau_j) = \xi\, P_{q+1-n}(\tau)\ , \qquad (7.136)$$

where the normalization constant ξ is of no interest in our context. The integral (7.134) thus may be brought into the form

$$\boldsymbol{\alpha}(t_0+h) = \xi \left(\frac{h}{2}\right)^{2q+3-2n} \boldsymbol{X}_{q+1-n} \int\limits_{-1}^{+1} P^2_{q+1-n}(\tau')\ d\tau'\ , \qquad (7.137)$$

where use was made of the orthogonality properties

$$\int\limits_{-1}^{+1} P_{q+1-n}(\tau')\ P_l(\tau')\ d\tau' = 0\ , \quad l \neq q+1-n$$

of the Legendre polynomials. Assuming that the Legendre polynomials $L_i(\tau)$ are fully normalized, we may even write:

$$\boldsymbol{\alpha}(t_0 + h) = \xi \left(\frac{h}{2}\right)^{2q+3-2n} \boldsymbol{X}_{q+1-n} \; . \tag{7.138}$$

Observe that the excellent error property (7.138) can only be achieved for one time argument, namely for $t_0 + h$. For any other time argument $t_0 + \Delta t$, the error behavior of the numerical solution is given by eqn. (7.134). The improvement achieved at $t_0 + h$ by selecting the collocation epochs as the roots of the Legendre polynomial of degree $q{+}1{-}n$ w.r.t. every other selection are substantial. One achieves in essence a doubling of the error order!

7.6.5 Collocation Applied to Numerical Quadrature

So far, we only considered solutions of ordinary differential equation systems giving an approximating function as a result. If we apply the same principle to the numerical solution of integrals, this simply means that we want to find the *indefinite integral* or the *primitive function* $y(t)$ of a function $f(t)$ in an interval:

$$\int f(t')\; dt' = y(t) \, + \, C \; , \tag{7.139}$$

where, in our applications, the integration constant C is usually given by the value of $y(t)$ at an initial epoch t_0. By taking the first derivative of this equation we obtain the underlying differential equation:

$$\dot{y} = f(t) \; . \tag{7.140}$$

With this understanding it is sufficient for the solution of integrals to consider only scalar equations of first order, because the vectorial system of d equations of type

$$\dot{\boldsymbol{y}} = \boldsymbol{f}(t) \tag{7.141}$$

is actually decomposed into d separate scalar equations, and because for an equation of order $n > 1$ one may first calculate the solution function $\boldsymbol{y}^{(n-1)}$ and then integrate this approximating function $\boldsymbol{y}^{(n-1)}(t)$ $n-1$ times – which does not pose a problem when using polynomials as approximating functions.

The *definite integral* between the limits t_0 and $t_0 + h$ may now be simply defined as the difference of the values of the primitive function at $t = t_0 + h$ and $t = t_0$:

$$\int\limits_{t_0}^{t_0+h} f(t')\; dt' = y(t_0 + h) \, - \, y(t_0) \; . \tag{7.142}$$

With these introductory remarks the problem of numerical quadrature was considerably reduced. The system of condition equations may be transcribed from eqns. (7.125)

$$\begin{aligned}\sum_{l=1}^{q} \frac{1}{(l-1)!} \left(t_{0_j} - t_0^*\right)^{l-1} y_0^{(l)} &= f(t_{0_j}) \, , \quad j = 1, 2, \ldots, q \\ \sum_{l=1}^{q} \frac{\tau_j^{l-1}}{(l-1)!} \left(\frac{h}{2}\right)^{l-1} y_0^{(l)} &= f(t_{0_j}) \, , \quad j = 1, 2, \ldots, q \, ,\end{aligned} \tag{7.143}$$

where the origin of the development $t_0^* \overset{\text{def}}{=} t_0 + \frac{h}{2}$ (for numerical and for symmetry reasons) was selected as the midpoint of the interval $I_0 = [t_0, t_0 + h]$ and the relative time argument τ assuming the values $\tau(t_0) = -1$ and $\tau(t_0 + h) = +1$ is defined by the transformation

$$\tau \overset{\text{def}}{=} \frac{2(t - t_0^*)}{h} \; . \tag{7.144}$$

The above system of condition equations may be written in matrix form:

$$\mathbf{M} \boldsymbol{Y} = \boldsymbol{F} \; , \tag{7.145}$$

where

$\boldsymbol{Y}^T \overset{\text{def}}{=} \left(\dot{y}_0, \left(\frac{h}{2}\right) \ddot{y}_0, \ldots, \left(\frac{h}{2}\right)^{q-1} y_0^{(q)}\right)$,

$\mathbf{F}^T \overset{\text{def}}{=} \left(f(t_{0_1}), f(t_{0_2}), \ldots, f(t_{0_q})\right)$, and

$\mathbf{M}$ is a matrix of dimension q with the general element

$\mathrm{M}_{jl} = \frac{\tau_j^{l-1}}{(l-1)!} \, , \; i = 1, 2, \ldots, q \, , \; j = 1, 2, \ldots, q \, .$

With the use of the relative time argument τ, the matrix $\mathbf{M}$ depends on the order q and on the specific selection of the time arguments τ_j , but not on the particular problem. This property allows it to compute the indefinite and the definite integral with very concise formulae.

We observe in particular that each coefficient $y_0^{(l)}$ may be written as a linear combination of the q function values $f(t_{0_j})$, $j = 1, 2, \ldots, q$, and that in consequence the definite integral between the limits t_0 and $t_0 + h$ also is a linear combination of the q function values $f(t_{0_j})$, $j = 1, 2, \ldots, q$. The result therefore may be written in the form:

$$\int_{t_0}^{t_0+h} f(t') \, dt' = y(t_0 + h) \; - \; y(t_0) = \frac{h}{2} \sum_{i=1}^{q} W_i \, f\big(t(\tau_i)\big) \; , \tag{7.146}$$

where for every subdivision of the interval I_0 the following error-order holds:

$$\int_{t_0}^{t_0+h} f(t')\, dt' = y(t_0+h) - y(t_0) = \frac{h}{2}\sum_{i=1}^{q} W_i\, f(t'_i) + O\left(h^{q+1}\right) \ . \quad (7.147)$$

If the relative time arguments are identified with the roots of the Legendre polynomial of order q, the corresponding equation reads as

$$\int_{t_0}^{t_0+h} f(t')\, dt' = y(t_0+h) - y(t_0) = \frac{h}{2}\sum_{i=1}^{q} W_i\, f\big(t(\tau_i)\big) + O\left(h^{2q+1}\right) \ , \quad (7.148)$$

which is of course much better than what can be achieved by an "arbitrary" (e.g., an equidistant) selection of the epochs τ_i . These particular methods of numerical quadrature are called *Gaussian quadrature formulae.* They are truly outstanding for the evaluation of definite integrals. There is no need to prove the relationship (7.148): the equation is a consequence of the more general result (7.138).

Table 7.4 gives the roots of the Legendre polynomials and the weights associated with the corresponding quadrature formulae up to order $q = 10$. Due to symmetry properties of the Legendre polynomials it is only necessary to include the roots and weights in the right-hand side $[0,1]$ of the transformed integration interval $I'_0 = [-1,+1]$. The other roots are defined by $\tau_i = -\tau_{q+1-i}$, and the associated weights by $W_i = W_{q+1-i}$, $i = 1, 2, \ldots, q$. An integration routine, up to order $q = 30$ is used in the included program package to solve integrals, where the limiting order is $q_{\max} = 30$ – implying that the error of the approximation is of the order $O(h^{61})$.

Not all (but most) problems related to numerical quadrature may be solved using the Gaussian quadrature formulae. Exceptions are perhaps the quadrature of empirically given functions (where it may be much simpler to derive a table of equally spaced function values or of functions which are provided in tabular form). One can counter this argument by the remark that it is always possible to deduce function values at the epochs τ_i by interpolation. One should, however, keep in mind that eqn. (7.148) is based on the assumption that the function values are free of errors (except rounding errors limiting the accuracy).

The result (7.138) assumes that the initial value problem is associated to a linear differential equation system (which formally includes the case of an integral). The question is whether the same subdivision of the integration intervals I_k could also be used for the general case. This is possible in principle. The success is however limited by the fact that the function values at the epochs τ_i need to be known with an accuracy of the order of h^{2q+1} – which would ask for a rather large number of iterations in each subinterval. Also, the accumulation of rounding errors would have to be considered.

Gaussian integration procedures may also be used to solve improper integrals. We refer to [88] and to [15] for more information.

Table 7.4. Roots τ_i of the Legendre polynomials and weights W_i of the Gaussian integration procedure

Order q	Index i	Root τ_i	Weight W_i
1	1	0.0000000000000000	2.0000000000000000
2	2	0.5773502691896257	1.0000000000000000
3	2	0.0000000000000000	0.8888888888888889
	3	0.7745966692414834	0.5555555555555555
4	3	0.3399810435848563	0.6521451548625462
	4	0.8611363115940526	0.3478548451374539
5	3	0.0000000000000000	0.5688888888888889
	4	0.5384693101056831	0.4786286704993665
	5	0.9061798459386640	0.2369268850561891
6	4	0.2386191860831969	0.4679139345726910
	5	0.6612093864662645	0.3607615730481387
	6	0.9324695142031521	0.1713244923791703
7	4	0.0000000000000000	0.4179591836734694
	5	0.4058451513773972	0.3818300505051189
	6	0.7415311855993945	0.2797053914892768
	7	0.9491079123427586	0.1294849661688700
8	5	0.1834346424956498	0.3626837833783619
	6	0.5255324099163290	0.3137066458778873
	7	0.7966664774136267	0.2223810344533744
	8	0.9602898564975362	0.1012285362903758
9	5	0.0000000000000000	0.3302393550012598
	6	0.3242534234038089	0.3123470770400028
	7	0.6133714327005904	0.2606106964029355
	8	0.8360311073266358	0.1806481606948573
	9	0.9681602395076261	0.0812743883615744
10	6	0.1488743389816312	0.2955242247147529
	7	0.4333953941292472	0.2692667193099962
	8	0.6794095682990244	0.2190863625159821
	9	0.8650633666889845	0.1494513491505805
	10	0.9739065285171717	0.0666713443086881

Our treatment of numerical quadrature has shown that the famous Gaussian quadrature formulae are nothing but a special case of collocation methods, where the "speciality" consists of the selection of the collocation epochs τ_j . The approach was rewarding insofar as we saw that the solutions of linear differential equation systems have similar properties as the solutions of integrals.

Let us conclude the topic of numerical quadrature by the remark that the Gaussian integration procedures may be introduced in many different ways.

The analysis by Z. Kopal in [63], discussing an algebraic, a geometrical, and an analytical approach is an excellent reference.

7.6.6 Collocation: Examples

Program LINEAR (see Chapter II- 6 of Part III) demonstrates the capabilities of collocation techniques when applied to linear differential equations (or systems of equations) and to integrals. LINEAR is based on a Fortran-77 subroutine, which is capable of generating (almost) the most general particular solution that can be associated with a linear differential equation system of order $n > 0$. The order n, the approximation order q, and the dimension d of the system are input variables of this subroutine.

The program LINEAR allows it to solve nine linear initial or boundary value problems. Two of the problems are integrals, some of them are identical from the mathematical point of view. The program produces a general output file, an error file, and a tabular output file. The general output file contains the coefficients of the numerical solution and all components of the solution vector. The tabular output file contains three columns, namely the independent argument in the first column, the first component of the numerically integrated solution in the second, and the associated error in the third column. More information may be found in Chapter II- 6 of Part III. The problems addressed by program LINEAR, in the order as they are proposed in the primary menu of program LINEAR (see Chapter II- 6 of Part III), are:

1. *Exponential function:*

$$\begin{aligned} \dot{y} &= -y \\ y(0) &= 1 \end{aligned} \tag{7.149}$$

   with the true solution

$$y(t) = e^{-t} \ . \tag{7.150}$$

2. *Harmonic Oscillator:*

$$\begin{aligned} \ddot{y} &= -y \\ y(0) &= 1 \\ \dot{y}(0) &= 0 \end{aligned} \tag{7.151}$$

   with the true solution

$$y(t) = \cos t \ . \tag{7.152}$$

3. *Harmonic Oscillator:* Same as problem (7.151), but solved as first order system.

4. *Bessel's differential equation* (in a somewhat unusual explicit version):

$$\begin{aligned} \ddot{y} &= -\frac{t^2 - n_b^2}{t^2}\, y - \frac{1}{t}\, \dot{y} \\ n_b = 0: \quad y(0) &= 1 \\ \dot{y}(0) &= 0 \\ n_b > 0: \quad y(0) &= 0 \\ y^{(n_b)}(0) &= \frac{1}{2^{n_b}} \end{aligned} \tag{7.153}$$

with the true solution

$$y(t) = J_{n_b}(t) \ , \tag{7.154}$$

where $J_{n_b}(t)$ is the Bessel function (of the first kind) with pointer n_b (named after Friedrich Wilhelm Bessel (1784–1846)).

5. *Bessel's differential equation* (same as above problem), but solved as first order system.

6. $$\begin{aligned} \dot{y} \quad &= \sin t \ y \\ y(0) &= e^{-1} \end{aligned} \tag{7.155}$$

with the exact solution

$$y(t) = e^{-\cos t} \ . \tag{7.156}$$

7. *Legendre's differential equation*:

$$\begin{aligned} \ddot{y} \qquad &= \frac{2\,t}{1-t^2}\, \dot{y} - \frac{n_l\,(n_l - 1)}{1-t^2}\, y \\ y(-1) &= (-1)^{n_l} \\ \dot{y}(1) \quad &= +1 \end{aligned} \tag{7.157}$$

with the Legendre polynomial $L_{n_l}(t)$ of degree n_l (for integer values $n_l \geq 0$) as true solution.

8. $$\begin{aligned} \dot{y} \quad &= \cos t \\ y(0) &= 0 \end{aligned} \tag{7.158}$$

with the solution

$$y(t) = \int\limits_0^t \cos t' \ dt' = \sin t \ . \tag{7.159}$$

9. *Special problem posed by Z. Kopal*: The problem is taken from Z. Kopal [63], where it is used to demonstrate the capabilities of methods for repeated integrations

$$\begin{aligned} \ddot{y} &= \frac{1+3\,t+2\,t^3}{(1+t^2)\sqrt{1+t^2}}\cos\sqrt{1+t^2} \;-\; \frac{t^2\,(1+t)}{1+t^2}\sin\sqrt{1+t^2} \\ y(-1) &= 1 \\ y'(0) &= 0\;. \end{aligned} \tag{7.160}$$

The solution is

$$y(t) = 1 + (1+t)\left[\sin\sqrt{1+t^2} \;-\; \sin(1)\right]\;. \tag{7.161}$$

It cannot be the intention to deal with all possible questions that might be answered with program LINEAR. The following tests are meant to address a few key issues and to stimulate further investigations. The key issues are:

- Establishment of the impact of the distribution of collocation epochs. Program LINEAR allows to select three options (see Chapter II- 6 of Part III):
  1. Equidistant distribution in the integration interval I.
  2. For the integration order q the collocation epochs are the roots of the Chebyshev polynomial (named after Pafnutij Lwowitsch Chebyshev (1821–1894)) of degree $q+1-n$ in the interval $I' = [-1,+1]$ (transformation of the original interval $I = [t_0, t_0+h]$).
  3. For the integration order q the collocation epochs are the roots of the Legendre polynomial of degree $q+1-n$ in the interval $I' = [-1,+1]$ (transformation of the original interval $I = [t_0, t_0+h]$).
- The interval lengths typically may be much longer than in the case of non-linear differential equation systems.
- High integration orders (up to order $q \approx 30$) are allowed by program LINEAR (and may make sense).
- Integrable singularities (as they are encountered, e.g., in problems (7.153), (7.157), and (7.160)) do not matter, as long as none of the collocation epochs is identical with the argument at which the singularity occurs.
- The accuracy of the solution at the right interval boundary $t = t_0 + h$ is best, if the collocation epochs are defined as the roots of the Legendre polynomials (confirmation of result (7.138)).

Let us start by solving problem (7.149) using an approximation order of $q = 20$ in the interval $I = [0, 10]$ and all three epoch selections (equidistant, Chebyshev, and Legendre). The high order and the length of the integration interval impressively confirm the above statements concerning the interval length and the order of the approximation: Linear and non-linear problems really are problems of different categories. The approximation error

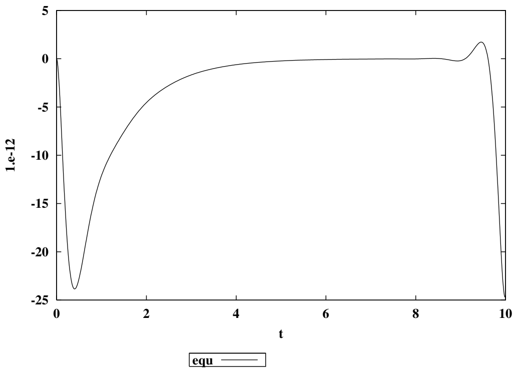


Fig. 7.19. Integration error in problem 1 using equidistant collocation epochs, $q = 20$

for equidistant collocation epochs, in units of 10^{-12}, is shown in Figure 7.19. The result is satisfactory throughout the interval. One should, however, keep in mind that for $t = 10$ one has $y(t) = e^{-10} \approx 4.5 \cdot 10^{-5}$, which means that the relative accuracy at the right interval boundary is only $\Delta y/y \approx 5.5 \cdot 10^{-5}$.

The results are much better, if the collocation epochs are chosen as the roots of the Chebyshev or the Legendre polynomials. Figure 7.20 shows that near the interval boundaries the errors are about a factor of 150 smaller than in the first case! The improvement is striking and the conclusion convincing: If linear equations or integrals are solved numerically, the spacing between collocation epochs should not be equidistant, but either be defined by the spacing between the roots of the Chebyshev or the Legendre polynomials. Figure 7.20 also shows, that on the average, over the entire integration interval, the quality of the approximations is similar in both cases (Chebyshev and Legendre), the latter case being perhaps a factor of 1.1 to 1.2 better. The result is (as expected) different at the right interval boundary (documented by the blow-up of the results in the interval $[9.5, 10]$). The error at the right interval boundary is a factor of $\frac{9.4 \cdot 10^{-18}}{9.1 \cdot 10^{-14}} \approx 10^{-4}$ smaller if the collocation epochs are selected as the roots of the Legendre polynomial of degree $q + 1 - n$ and not as the roots of the Chebyshev polynomial of the same degree. This characteristic is completely irrelevant, if the purpose of integration is the generation of an approximating solution within the entire integration interval. It does, however, matter, if the solution vector shall be propagated over many intervals. Unnecessary to say that program LINEAR could be easily generalized to serve this purpose.

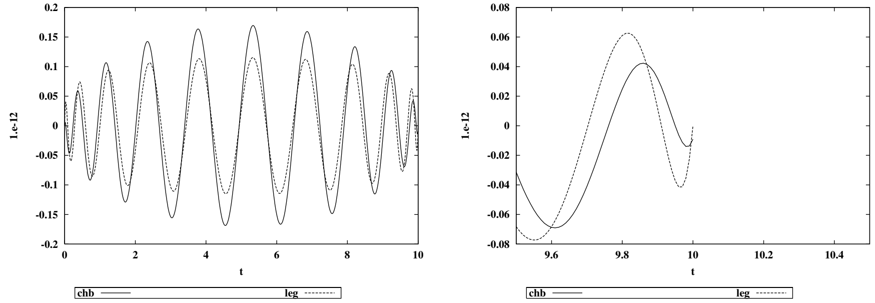


Fig. 7.20. Integration error in problem 1 using collocation epochs as roots of Chebyshev and Legendre polynomials $q = 20$ (left: entire interval, right: error behavior near the right boundary)

The above results based on the solution of problem (7.149) may be easily confirmed using the other examples. It is, e.g., instructive to perform the same investigation as above using problem (7.151) and the same interval length of $h = 10$, corresponding to about 1.6 periods and the same order $q = 20$ of the approximation. The accuracy (absolute and relative) of the solution is of the order of $8.6 \cdot 10^{-8}$ for the equidistant, of $3.2 \cdot 10^{-12}$ for the Chebyshev, and of $6.7 \cdot 10^{-15}$ for the Legendre selection of collocation epochs. We might view the solution $\cos t$ as one component of a solution vector of a circular orbit. In the above computation we have achieved a solution with a relative accuracy of $\sim 10^{-15}$ after 1.6 revolutions with just $q+1-n = 20+1-2 = 19$ function evaluations! This has to be compared to the results of the two-body problem for circular orbits, where the same integration method (collocation) required between 300 and 1000 function evaluations per revolution. These facts underline that exploiting the linearity of differential equation systems may very well reduce the computational effort.

Let us briefly discuss the differential equation (7.157) for the Legendre polynomials. The problem is posed as a classical boundary value problem. Moreover, integrable singularities occur at $t = \pm 1$. This is why the equidistant distribution of collocation epochs cannot be used (otherwise the first and the last grid-point would coincide with $t = \pm 1$). An error message is produced if one tries to use this option, and processing is terminated. Also, as we know that the solution will be a polynomial, the order of the method is constrained to $q \overset{\text{def}}{=} n_l$, where n_l is the degree of the Legendre polynomial considered. If one solve this problem with either the Chebyshev or the Legendre distribution of collocation epochs (using equidistant spacing is not allowed), you will not observe the accuracy differences due to the collocation epochs mentioned above: If the true solution of a differential equation is a polynomial of degree q, this solution will be reproduced by any of the methods (equidistant, Chebyshev, Legendre), apart from the rounding errors (assuming that there are no problems with singularities involved).

Problem (7.153) (Bessel's differential equation) represents a conventional initial value problem for $n_b = 0$ for a differential equation of order $n = 2$, by providing the solution vector and its first derivative at $t = 0$. For higher orders $n_b > 0$, the solution vector and a higher (than the first) derivative are provided at $t = 0$. The program LINEAR is obviously capable of handling such generalized initial value problems – where $n_b > q$ would of course lead to a disaster.

Using Bessel's differential equation we try to answer the question whether it is better to decompose a higher-order system into one of first order or not, from the accuracy point of view. For the pointer $n_b = 0$ both solution methods can be used (options "BESSEL" for solving directly the second order system, "BESSEL(1ST_ORDER)" for solving the same problem after decomposition into a system of first order). Figure 7.21 reminds us of the Bessel-function $J_0(t)$ in the interval $I = [0, 10]$ (generated with a collocation procedure of order $q = 20$).

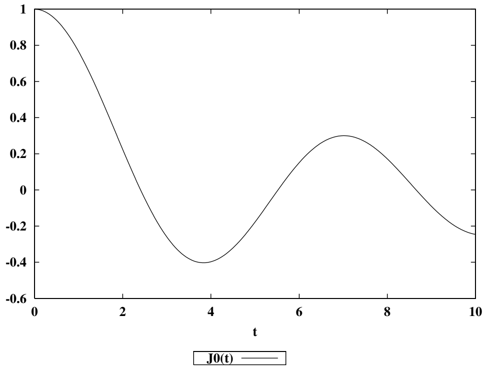


Fig. 7.21. Bessel function using $q = 20$ and Legendre distribution of collocation epochs

Figure 7.22 contains the errors for case 1, a direct integration of the second-order equation with order $q = 20$, for case 2, an integration of the first-order system with order $q = 20$, and for case 3, an integration of the second-order equation with an approximation order of $q = 21$. (The collocation epochs were chosen as the roots of the Legendre polynomials). It must be said that the three cases are not completely equivalent from the theoretical point of view, and it is somewhat arbitrary which of the three cases should be considered as equivalent. One may argue that cases 1 and 2 are equivalent, because the solution vector $y(t) = J_0(t)$ is represented by a polynomial of

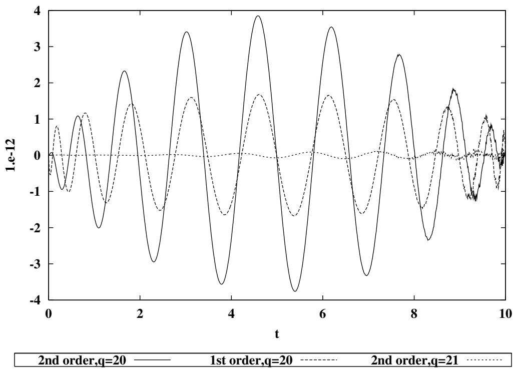


Fig. 7.22. Errors of Bessel function $J_0(t)$ using orders $q = 20, 21$, 2nd and 1st order differential equations, and Legendre distribution of collocation epochs

the same degree in the two cases. One may also argue that cases 3 and 2 are equivalent because the first derivative of $J_0(t)$ is represented by polynomials of the same degree, and, what is more important, because the same number of collocation epochs were used in the two cases. Be this as it may: From Figure 7.22 we conclude that one should always try to solve directly the higher-order equation (or system of equations) when using collocation methods. This statement is strongly supported by the fact, that the dimension of the linear system of equations to be solved is roughly doubled, if a second order equation is replaced by a first order system.

In order to demonstrate that our algorithm may also be used to solve rather tricky questions involving repeated integrals – and in order to prove that no dedicated methods are required for this purpose – we solve Kopal's example (7.160) in the interval $I = [-1, 0]$ using a collocation method of order $q = 20$. Figure 7.23 shows the integrated function, Figure 7.24 shows the associated error. The interested reader may wish to compare these accuracies with the results reported in [63].

7.7 Error Propagation

The two types of errors to be considered in numerical integration, namely

- the *approximation errors*, due to the truncation of the local approximating functions, and
- the *rounding errors*, due to the fact that each number only is represented by a finite number of digits,

were already identified in section 7.5.4.

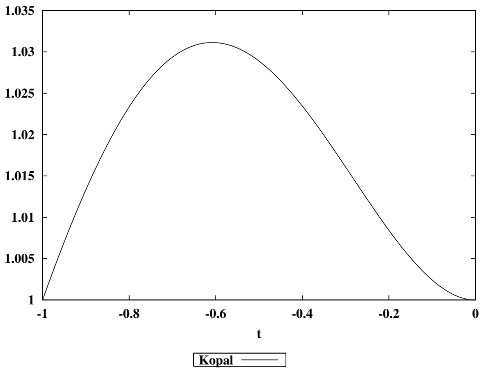


Fig. 7.23. Kopal's example using $q = 20$ and Legendre distribution of collocation epochs

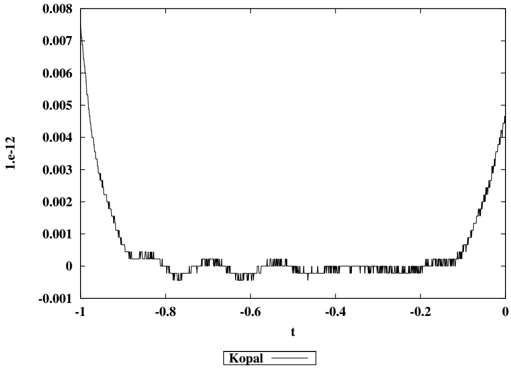


Fig. 7.24. Error of numerical solution of Kopal's example using $q = 20$ and Legendre distribution of Collocation epochs

The numerical examples in sections 7.5.4 and 7.5.5 gave a first impression of the accumulation of these errors when integrating two-body orbits. In this chapter we derive (some of) the laws of error propagation. Many of these laws are problem independent, others are only applicable to the (pure and perturbed) two-body motion.

In order to understand the accumulation of rounding errors the essential elements of computer-based algebra are reviewed briefly in section 7.7.1. Section 7.7.2 deals with rounding errors and their statistical prediction. Two-body orbits illustrate the theoretical developments. We first define the rounding errors introduced into the set of initial values referring to epoch t_k. From these local errors one may calculate the resulting changes of the first integrals, i.e., of the orbital elements. The error of the semi-major axis a is provided in particular. The accumulated rounding error in one of the first integrals at epoch t_N, due to the errors introduced at the epochs t_k, $k = 1, 2, \ldots, N$, may then be easily calculated as the plain sum of the epoch-specific errors in the first integral considered. The error propagation in the semi-major axis is of particular importance, because an error in the semi-major axis a, introduced at epoch t_k, induces an error in the mean motion, therefore also in the mean anomaly $\sigma(t)$ for the time $t \geq t_k$. The accumulated error in the mean anomaly $\sigma(t)$ due to the epoch-specific errors in a is derived towards the end of this section.

The approximation error and its accumulation are studied in section 7.7.3. The local error function referring to t_k is provided for the two-body motion for collocation and multistep methods. Two case studies illustrate the problems. Based on the results of sections 7.7.2 and 7.7.3 a rule of thumb for the optimal stepsize for orbits with small eccentricities is derived in section 7.7.4.

Some results of the error propagation theory developed up to section 7.7.4 are strictly valid only for the two-body motion. They may, however, also be used to describe approximately the error propagation of the perturbed two-body motion, provided the perturbations are small compared to the main term, and provided the integration interval is not excessively long. Should one or both of these assumptions be wrong, the error propagation theory must be put on a more general basis. The variational equations, as introduced in Chapter 5, are the essential tools in this case. These advanced aspects of error propagation are briefly sketched in the concluding section 7.7.5 of this chapter.

7.7.1 Rounding Errors in Digital Computers

In digital computers *rounding* of the result of an arithmetic operation is a part of the floating-point arithmetic environment. Each real number x is approximated by a rational number r, called floating-point number. Each floating-point number r is in turn represented by

- the absolute value $|r|$ of r, defined by an integer mantissa M consisting of a fixed number m of digits d_i in the system of numbers with base b,
- the sign s of r, and
- the integer exponent e in the system of numbers with base b.

Each floating-point number r may be brought into the form

$$r = s\, M\, b^{\,e-m+1} = s\, \{d_1 d_2 \ldots d_m\}\, b^{\,e-m+1} \ , \tag{7.162}$$

where d_1 is the most significant, d_m the least significant digit of the mantissa M. We may think of s as an integer assuming either the value $s = +1$ for

$r \geq 0$ or $s = -1$ for $r < 0$. Internally, the sign occupies only one "bit", the basic storage unit. A bit may assume two states – exactly what is needed to define a sign or one digit in the binary system (base $b = 2$).

The floating-point number is called "normalized", if the most significant digit $d_1 \neq 0$. Due to the representation (7.162), a normalized number r with exponent $e = 0$ has a value

$$1 \leq |r| < b \; . \tag{7.163}$$

The precision of a particular floating-point arithmetic environment is referred to as the smallest number ε_m, which, when added to the number 1, gives a floating-point number $\tilde{r} \neq r$. Roughly speaking, ε_m may be identified with the numerical value of the least significant digit d_m of a number r with $|r| \approx 1$ and an exponent $e_r = -1$:

$$\varepsilon_m = 1 \cdot b^{-m+1} \; . \tag{7.164}$$

The program system described in Part III is based on FORTRAN double precision in an environment, for which

$$b = 2 \quad \text{and} \quad m = 53 \; , \tag{7.165}$$

which implies that

$$\varepsilon_m = 2^{-52} = 2.22 \cdot 10^{-16} \; . \tag{7.166}$$

Different values may result if the program system is compiled using a different floating-point environment. For more information concerning floating-point representation and machine-specific questions we refer to [88].

Let r be the floating-point result of an arithmetic operation, the exact value of which is x. The result of this operation usually is a floating point number with more than m digits. Consequently, this internal result has to be rounded to the nearest floating-point number with m digits. The statistical characteristics of the rounding operation are the expectation value $E(r)$ and the variance $\text{var}(r)$ of the result. When rounding the number x it is assumed that the expectation value $E(r)$ of the rounded number r equals the true value:

$$E(r) = x \; . \tag{7.167}$$

From now on we assume that perfect rounding is implemented. If the intermediary result (before the rounding operation) were available with an infinite number of digits, the variance $\text{var}(r)$, expressed in units of the least significant digit of the result x, is given by:

$$\text{var}(r) = \text{var}(r - x) = E\big((r - x)^2\big) \int_{-0.5}^{0.5} x'^2 \, dx' = \frac{1}{12} \; . \tag{7.168}$$

The situation is more complicated if the intermediary result (before rounding) is only available with $m + 1$ digit (e.g., in the binary system of numbers). In the case of numerical integration we may, however, usually assume that more digits are available – at least where the essential arithmetic operations are concerned. Therefore we assume that eqns. (7.167) and (7.168) hold. Variance and expectation value of a floating point operation are thus assumed to be given by:

$$E(r) = x \text{ and } \mathrm{var}(r - x) = E\big((r - x)^2\big) = \frac{\varepsilon_m^2}{12}\, b^{2e_r} \ . \tag{7.169}$$

Equations (7.169) are the basis for the statistical treatment of rounding errors.

Rounding is of vital importance, but it is not necessarily implemented in all floating-point environments. In Chapter 20, Press et al. [88], provide so-called "less numerical algorithms" capable of extracting the "essentials" of a floating-point arithmetics environment.

As a side-remark we mention that the author, when writing his Ph.D. thesis, was exposed to a computer environment where numbers were truncated and not rounded. The consequences were rather dramatic: Because in the case of truncation $E(|r|) < |x|$ for each arithmetic operation (where x is the true, r the floating-point result), any numerically integrated orbit inevitably was shrinking (similar results would be obtained with correct rounding in the presence of a drag (!)). For more information consult [9].

If rounding is done properly, the expectation value of the numerical approximation equals its true value $E(r) = x$. The actual error of a numerically integrated trajectory is therefore governed by the variance of the accumulated errors and not by their expectation values.

7.7.2 Propagation of Rounding Errors

Local Rounding Errors in the Initial Values. Assuming that eqns. (7.169) hold it is *in principle* a straightforward procedure to calculate the resulting rounding error even for complicated algorithms. Keeping in mind the complexity of the algorithms for numerical integration, such a correct and straightforward approach is, however, hopelessly complicated, *in practice.* Thanks to the hierarchy of arithmetic operations we may fortunately reduce this complexity by assuming that each component of the new initial values calculated at the boundary t_k of the subinterval I_{k-1} contains exactly one rounding error. These rounding errors are defined by:

$$\tilde{\boldsymbol{y}}^{(i)}(t_k) = \boldsymbol{y}^{(i)}(t_k) + \boldsymbol{\varepsilon}_{ki} \ , \quad i = 0, 1, \ldots, n-1 \ , \tag{7.170}$$

where $\tilde{\boldsymbol{y}}^{(i)}(t_k)$ is the rounded floating-point result, $\boldsymbol{y}^{(i)}(t_k)$ the true result, and $\boldsymbol{\varepsilon}_{ki}$ the column matrix of the rounding errors of the initial value matrix $\boldsymbol{y}^{(i)}(t_k)$. Approximation errors are *not* considered (assumed to be zero) here.

We assume, in other words, that the rounding errors introduced at t_k are merely due to one arithmetic operation "new initial value at t_k = previous initial value at t_{k-1} plus increment". The old initial values and the increment are considered as errorfree in this context. Moreover we assume that different rounding errors are independent, leading to a diagonal variance-covariance matrix (containing the expectation values of the products $\varepsilon_{ki_l}\,\varepsilon_{ki_m}$, $l, m = 1, 2, \ldots, d$; the variances of the errors ε_{ki_l} reside on the diagonal of this matrix, the so-called covariances, i.e., the expectation values $E(\varepsilon_{ki_l}\varepsilon_{ki_m})$, $l \neq m$, are the off-diagonal elements):

$$\begin{aligned} E(\boldsymbol{\varepsilon}_{ki}) &= \mathbf{0}\ , \quad i = 0, 1, \ldots, n-1 \\ \mathrm{cov}(\boldsymbol{\varepsilon}_{ki}) &= \frac{\varepsilon_m^2}{12} \begin{pmatrix} b^{2e_{ki_1}} & 0 & 0 & \ldots & 0 \\ 0 & b^{2e_{ki_2}} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & b^{2e_{ki_d}} \end{pmatrix}, \quad i = 0, 1, \ldots, n-1\ , \end{aligned} \tag{7.171}$$

where e_{ki_l} is the exponent in the binary system $b = 2$ of the component l of derivative i of the new initial value $\boldsymbol{y}(t_k)$.

The model (7.170, 7.171) is general in the sense that it may be used for any initial value problem, not just in Celestial Mechanics. Equations (7.170, 7.171) assume that – from the point of view of the rounding errors – each algorithm is equivalent to the Euler algorithm (7.11, 7.12) and that the rounding errors in the increments ("new minus old" initial values) are negligible (and that the rounding takes place when adding the small increment to the "old" initial values). Due to the neglect of rounding errors in the increments the model is expected to be a fair approximation of the real situation for small stepsizes h_k.

If the solution vector $\boldsymbol{y}(t)$ is a (quasi-)periodic function of time, as it is often the case in Celestial Mechanics, the assumptions (7.171) may be further simplified. We may in essence forget about the component-specific exponents and use the approximation:

$$\begin{aligned} E(\boldsymbol{\varepsilon}_{ki}) &= \mathbf{0}\ , \quad i = 0, 1, \ldots, n-1 \\ \mathrm{var}(\boldsymbol{\varepsilon}_{ki}) &= \frac{\varepsilon_m^2}{12}\, b^{2e_{i,\max}}\, \mathbf{E}\ , \quad i = 0, 1, \ldots, n-1\ , \end{aligned} \tag{7.172}$$

where $e_{i,\max}$ is the maximum exponent assumed by the components of the derivatives i of the solution vector in its quasi-periodic movement in time t and $\mathbf{E}$ is the unit matrix of dimension d.

From now on we are going to use the simplified model (7.170, 7.172) for our developments.

Local Rounding Errors in First Integrals. Using the model (7.170, 7.172) for the errors introduced into the initial values at t_k, the error of any function of these initial values may be easily computed. It is in particular possible to compute the errors in the first integrals of motion (if they are known).

Let us apply the above results to the integration of the two-body problem. Let us denote the change in the semi-major axis due to the rounding errors introduced at t_k by δa_k. This error in the semi-major axis a of the two-body motion is calculated as follows (compare eqn. (4.20)):

$$\delta a_k = \frac{2\,a^2}{\mu} \left\{ \frac{\mu}{r_k^3}\, \boldsymbol{r}(t_k) \cdot \boldsymbol{\varepsilon}_{k0} \ + \dot{\boldsymbol{r}}(t_k) \cdot \boldsymbol{\varepsilon}_{k1} \right\} \ , \tag{7.173}$$

where $r_k = |\boldsymbol{r}(t_k)|$. Using the error model (7.170, 7.172) we obtain immediately

$$E(\delta a_k) = 0 \ , \tag{7.174}$$

because the expectation value of a linear combination of independent random variables is given by the same linear combination of the individual expectation values, and

$$\mathrm{var}(\delta a_k) = \frac{\varepsilon_m^2}{12}\, \frac{4\,a^4}{\mu^2} \left(\frac{\mu^2}{r_k^4}\, b^{2e_{0,\max}} \ + \dot{\boldsymbol{r}}_k^2\, b^{2e_{1,\max}} \right) \ , \tag{7.175}$$

because the variance of a linear combination of independent random variables is given by the same linear combination of the individual variances. For orbits with small eccentricities the approximations $r_k \approx a$ and $\dot{\boldsymbol{r}}_k^2 \approx n^2\,a^2$ may be used. Using in addition the relationship $n^2\,a^3 = \mu$ of the two-body problem, the above expression may be approximated by

$$\mathrm{var}(\delta a_k) = \frac{\varepsilon_m^2}{3} \left(b^{2e_{0,\max}} \ + \frac{1}{n^2}\, b^{2e_{1,\max}} \right) \approx \frac{2\,\varepsilon_m^2}{3}\, b^{2e_{0,\max}} \ . \tag{7.176}$$

The latter approximation is justified by the fact that $|\dot{\boldsymbol{r}}| = na$ for circular orbits, meaning that the errors in the velocity components, which originally were much smaller than the errors in the coordinates, are blown up to the same size as those in the coordinates. Observe, that in this approximation the variances do no longer depend on t_k and that the index k might be left out.

Accumulation of Rounding Errors in First Integrals. As the first integrals are constants in time, the accumulated error after N steps (subintervals) is simply computed as the sum of the local rounding errors introduced in each step. Using the approximation (7.176) for the local error in a we obtain the accumulated rounding error at t_N as the plain sum of the local rounding errors:

$$\delta a(t_N) = \sum_{k=1}^{N} \delta a_k \ . \tag{7.177}$$

The expectation value and the variance of this quantity is:

$$E\big(\delta a(t_N)\big) = 0 \tag{7.178}$$

and

$$\operatorname{var}\big(a(t_N)\big) = \frac{2\,\varepsilon_m^2}{3} \sum_{k=1}^{N} b^{2e_{0,\max}} = \frac{2}{3}\,\varepsilon_m^2\, b^{2e_{0,\max}}\, N \ . \tag{7.179}$$

Consequently, the mean error $\sigma\big(a(t_N)\big)$ in the semi-major axis due to the accumulation of the rounding errors may be approximated by

$$\sqrt{\operatorname{var}\big(a(t_N)\big)} = \sqrt{\tfrac{2}{3}}\,\varepsilon_m\, b^{e_{0,\max}}\, \sqrt{N} \ . \tag{7.180}$$

The result derived for the semi-major axis may be generalized: The accumulated rounding error of a first integral is always equivalent to the error of a sum of numbers, which in turn is equivalent to the error of a definite integral.

Let us illustrate this theory of rounding errors by an example. The orbit of Jupiter, with a semi-major axis of $a \approx 5.208$ AU (this implies $e_{0,\mathrm{max}} = 2$ in the binary system) and an eccentricity of $e \approx 0.048$, was numerically integrated over one million years (extending the integration interval of previous tests by a factor of one thousand). The integration was performed with program PLASYS using a multistep method of order $q = 14$ with stepsizes $h = 25$ days and $h = 30$ days, implying that 14.6 million integration steps had to be performed when using the stepsize of $h = 25$ days, 12.2 million steps when using a stepsize of $h = 30$ days.

Figure 7.25 shows the development of the semi-major axis and the corresponding error limits $\pm\sqrt{\operatorname{var}\big(a(t_N)\big)}$, which, according to the theory of normally distributed random errors, should contain 67% of the actual errors. It is important that the actual errors are of the same order of magnitude as the statistical estimates. They lie (almost) entirely within the error limits given by the above theory. We may certainly conclude that rounding is performed in the floating-point environment used (tests of this kind actually would reveal serious rounding deficiencies).

The errors in all other integrals of motion (the orbital elements) show a similar behavior. It is remarkable that, despite the fact that the integration of the Newton-Euler equations of motion was performed in rectangular coordinates, the errors in the first integrals are the errors corresponding to a sum (integral) of random errors with identical variances (and expectation value 0).

Accumulation of Rounding Errors in Mean Anomaly. The accumulated error $\delta\sigma(t_N)$ in the mean anomaly $\sigma(t_N)$ at t_N may be computed as a weighted sum of the local errors δa_k, where the weights are $(t_N - t_k)$:

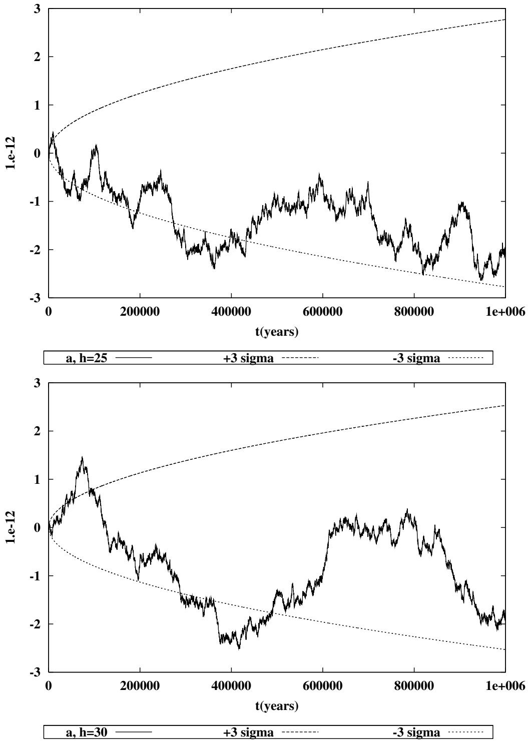


Fig. 7.25. Actual errors (in AU) in semi-major axis a and $\pm\sqrt{\mathrm{var}(a)}$ error limits for an integration of Jupiter over one million years in steps of 25 days (upper part) and 30 days with multistep method

$$\delta\sigma(t_N) = -\frac{3}{2}\,\frac{n}{a}\sum_{k=1}^{N}(t_N - t_k)\;\delta a_k\;. \tag{7.181}$$

The expectation value of $\delta\sigma(t_N)$ is of course zero

$$E\bigl(\delta\sigma(t_N)\bigr) = 0 \tag{7.182}$$

and the variance of this expression, for orbits with a small eccentricity, may be computed as

$$\mathrm{var}\bigl(\delta\sigma(t_N)\bigr) = \frac{9}{4}\,\frac{n^2}{a^2}\,\mathrm{var}(\delta a)\sum_{k=1}^{N}(t_N - t_k)^2\;. \tag{7.183}$$

If the stepsize is (close to) constant (as it is, e.g., the case for the multistep method used in the tests illustrated by Figure 7.25), we obtain

$$\sum_{k=1}^{N}(t_N - t_k)^2 = h^2\sum_{k=1}^{N}(N-k)^2 \approx h^2\,\frac{N^3}{3}\;. \tag{7.184}$$

Introducing the above equation into eqn. (7.183) and taking into account formula (7.176) gives the final result for the variance of the mean anomaly:

$$\mathrm{var}\big(\delta\sigma(t_N)\big) = \frac{9}{4}\,\frac{n^2}{a^2}\,\frac{2\,\varepsilon_m^2}{3}\,b^{2e_{0,\max}}\,h^2\,\frac{N^3}{3} = \frac{n^2\,h^2}{2\,a^2}\,\varepsilon_m^2\,b^{2e_{0,\max}}\,N^3\ . \tag{7.185}$$

The expected mean error of the anomaly after N integration steps of equal length is therefore given by the following equation:

$$\sqrt{\mathrm{var}\big(\delta\sigma(t_N)\big)} = \sqrt{\frac{1}{2}}\,\frac{n\,h}{a}\,\varepsilon_m\,b^{e_{0,\max}}\,\sqrt{N^3}\ . \tag{7.186}$$

Whereas the mean error (root of the variance) of a numerically integrated first integral was growing with a $\sqrt{N}$-law (N being the number of integration steps performed), the error in the mean anomaly is expected to grow according to a $\sqrt{N^3}$-law.

Figure 7.26 compares the actual errors in the argument of latitude u of the numerically integrated two-body orbit of Jupiter over one million years. The integration specifications are the same as those already mentioned in the context of Figure 7.25. For the small eccentricity of $e \approx 0.048$ of Jupiter's orbit the error in the argument of latitude u is expected to be close to the error in the mean anomaly. This is confirmed by Figure 7.26. It is satisfactory to see that the actual errors are well captured by the simple formula (7.186).

It is important to fix the orders of magnitude: whereas the errors in the first integrals, even after some 100000 revolutions, still are known to within a few parts in 10^{-12}, the argument of latitude (in degrees) may already contain errors of the order of 10^{-5} degrees, corresponding to few hundredths (10^{-2}) of an arcsecond. This, on the other hand, is sufficient for most studies one might wish to perform.

If algorithms for numerical integration are tested using the two-body problem, one is often not aware of the structure of error propagation as it was outlined here ($\sqrt{N}$-law for the integrals of motion, $N^{3/2}$-law for the mean anomaly). Usually, only the $N^{3/2}$-law for the argument of latitude, which also translates into a $N^{3/2}$-law for the error in the coordinates (and the velocities), is observed (if a power law is mentioned at all). The errors in the position vector have, for orbits with small eccentricities, the following characteristic structure:

$$\begin{aligned}\varDelta\mathbf{r}(t) &\approx \mathbf{R}_3(-\varOmega)\,\mathbf{R}_1(-i)\,\mathbf{R}_3(-\omega)\begin{pmatrix} a\,[\cos(\sigma+\delta\sigma)-\cos\sigma\,] \\ a\,[\sin(\sigma+\delta\sigma)-\sin\sigma\,] \\ 0 \end{pmatrix} \\ &\approx \mathbf{R}_3(-\varOmega)\,\mathbf{R}_1(-i)\,\mathbf{R}_3(-\omega)\begin{pmatrix} -\,a\,\delta\sigma\sin\sigma \\ +\,a\,\delta\sigma\cos\sigma \\ 0 \end{pmatrix}\ ,\end{aligned} \tag{7.187}$$

i.e., the errors in the coordinates show an oscillation with the revolution

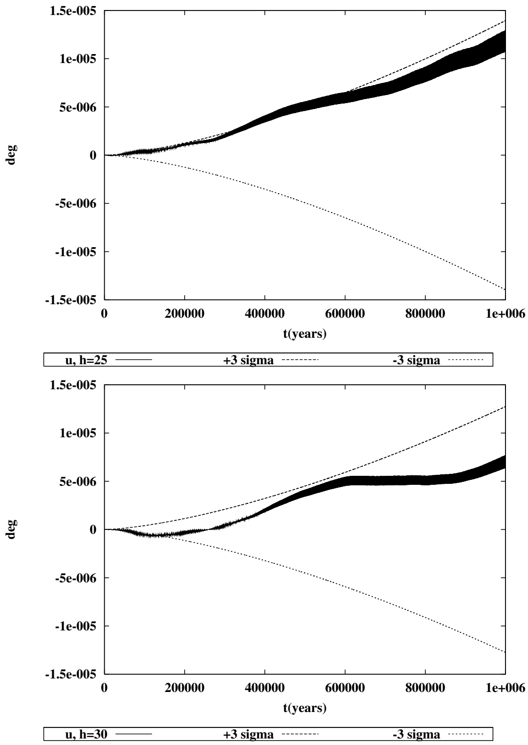


Fig. 7.26. Actual errors in argument of latitude u (in degrees) and $\pm\sqrt{\mathrm{var}(u)}$ error limits for an integration of Jupiter over one million years in steps of 25 days (upper part) and 30 days with multistep method

period as basic period and an amplitude growing with the $N^{3/2}$-law of formula (7.186).

The theory of the accumulation of rounding errors, as it developed in this paragraph, is based on the fundamental article [26] "On the accumulation of errors in numerical integration" by Dirk Brouwer (1902–1966) in 1937. Brouwer's laws were given for a special integration method and for the environment of "manual" computation (fixed number of decimal digits including leading zeros). This is why the coefficients in Brouwer's work and in our presentation differ somewhat. Formulae (7.180) and (7.186) are applicable to a broad class of integration methods. It is only required that the increments added to the old initial values are small (compared to the latter values) and that the rounding error of the increment itself is negligible. Formulae (7.180) and (7.186) describe the rounding laws for all single-step methods, in particular Taylor series, Runge-Kutta methods, collocation and those multistep methods equivalent to collocation methods.

7.7.3 Propagation of Approximation Errors

Approximation Errors: A First Case Study. Figures 7.25 and 7.26 might give the impression that it is possible to eliminate approximation errors completely in a numerical integration procedure. This is not true and we will see that eventually, if the integration interval is made long enough, approximation errors must become more important as an error source than rounding errors. There are two ways to make the approximation error visible, namely

1. by increasing the length of the integration interval – a method which might be pretty costly – or
2. by increasing the stepsize.

The latter option is used to introduce the problem: The orbit of Jupiter was numerically integrated as before (see Figures 7.25 and 7.26), but instead of using stepsizes of $h = 25$ or $h = 30$ days one of $h = 40$ days was used. Observe that a relatively modest change of the stepsize leads to a rather significant change in the error behavior. As before, the integration was performed with a multistep method (with one interpolation step).

The result in Figure 7.27 shows that the errors are much larger than expected by the error limits (expected for the rounding errors). Instead of a $\sqrt{N}$-law for the semi-major axis a and a $\sqrt{N^3}$-law for the argument of latitude u, a growth proportional to N in a and one proportional to N^2 in u are observed. The power-laws extracted from in Figure 7.27 support the above remark, that the approximation errors may be made visible (for any choice of h) by increasing the length of the integration interval.

The Local Error Function in Review. Equation (7.85) represents an approximation for the local error function of collocation methods. For multistep methods the corresponding formula follows from the representation (7.113). For convenience we include the explicit version of these error functions for both, the collocation and the multistep methods. As opposed to section 7.5.5, where we used the term of order q to get an estimate of the approximation error for the collocation method of order q for the purpose of error control, we have to use here the terms of order $q+1$ to assess the quality of a solution of order q. The error function then reads as:

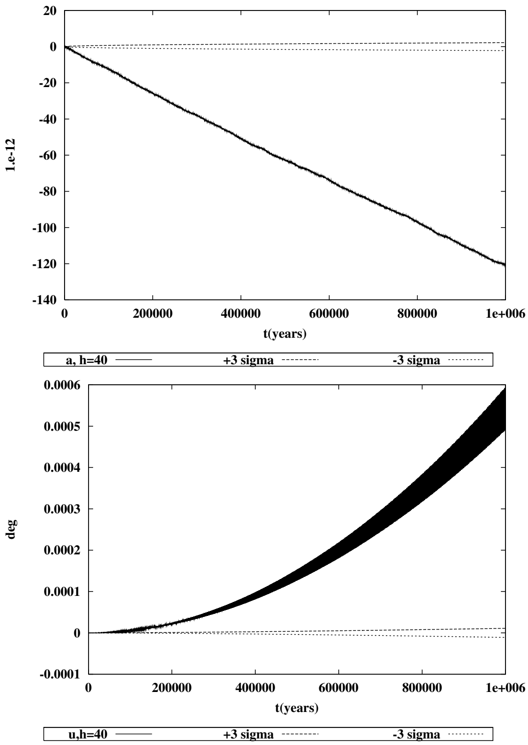


Fig. 7.27. Actual errors in semi-major axis a (upper part, in AU) and argument of latitude u (lower part) (including the corresponding error limits due to rounding) for an integration of Jupiter over one million years in steps of 40 days

$$\varepsilon_k^{(i)}(t) \overset{\text{def}}{=} \begin{cases} \text{collocation} \\ \left\{ \sum\limits_{l=n}^{q+1} \dfrac{(l-n)!}{(l-i)!} \, \tilde{\mathrm{M}}^{-1}_{l+1-n,j+1-n} \, \tau^{l-i} \right\} \left(\dfrac{h_k}{q+1-n} \right)^{q+1-i} \boldsymbol{y}_{k0}^{(q+1)} \ ; \\ \text{multistep} \\ \left\{ \sum\limits_{l=n}^{q+1} \dfrac{(l-n)!}{(l-i)!} \, \tilde{\mathrm{N}}^{-1}_{l+1-n,j+1-n} \, \tau^{l-i} \right\} h^{q+1-i} \, \boldsymbol{y}_{k0}^{(q+1)} \ ; \end{cases} \qquad (7.188)$$

$$i = 0, 1, \ldots, n-1 \ .$$

The structure of the error function formally proves the result (7.115), saying in essence that the spacing of the t_k must be made much narrower (by a factor of $(q-n)^{-1}$) in the case of the multistep methods (as compared to the collocation method). Only the formula for the pure extrapolation method is supplied in eqns. (7.188). The corresponding formula for the interpolation method is obtained by replacing $\tilde{\mathrm{N}}_{ik}$ by $\tilde{\mathrm{N}}_{\mathrm{int},ik}$ (see eqn. (7.106)).

The above error functions are approximations based on the assumption that all terms of the Taylor series higher than order $q+1$ are zero – what is of course not the case. For linear systems we showed that the correct error function solves the linear system of equations (7.127). It is an easy task to show that the same equation holds for general non-linear differential equation systems, *provided* the matrices $\mathbf{A}_i$ are interpreted as the Jacobian matrices of the function $\boldsymbol{f}$ with respect to the derivative of order i of the solution vector. Needless to say that the dependence of the true error function on the actual problem type is much more complicated than that of the simplified version (7.188), where the problem dependence is uniquely contained in the term $\boldsymbol{y}_{k0}^{(q+1)}$.

Be this as it may: we will assume from now on that the error function is given by eqn. (7.188). Observe that the terms in brackets in these equations may be interpreted as the derivatives of a scalar function $e_r(\tau)$ depending on the relative time argument τ (which is defined by (7.76) for collocation methods and by (7.110) for multistep methods). Making use of these scalar functions the error function (7.188) may be written as:

$$\boldsymbol{\varepsilon}_k^{(i)}(t) \overset{\text{def}}{=} \begin{cases} \dfrac{d^i}{d\tau^i}\left(e_{r,\text{colloc}}\right)\left(\dfrac{h_k}{q+1-n}\right)^{q+1-i} \boldsymbol{y}_{k0}^{(q+1)} \;; & \text{collocation} \\ \dfrac{d^i}{d\tau^i}\left(e_{r,\text{multi}}\right) h^{q+1-i}\, \boldsymbol{y}_{k0}^{(q+1)} \;; & \text{multistep} \\ & i = 0, 1, \ldots, n-1 \;. \end{cases} \tag{7.189}$$

Subsequently we want to analyze the error behavior of the unperturbed two-body motion. For this purpose we need to know the error function for differential equation systems of the order $n = 2$ at the right interval boundary. For collocation methods the right interval boundary corresponds to $\tau = q - n$, for multistep methods to $\tau = 1$. The results are contained in Table 7.5, where the error function is provided for the extrapolation and the interpolation method in the case of the multistep procedure. Table 7.5 is essential for properly understanding the error characteristics of collocation and multistep methods. Its implications are:

- The table underlines the equivalence of multistep and collocation methods from the point of view of the error function – provided the step ratio is defined by $h_{\text{colloc}} : h_{\text{multi}} = (q - n)$.
- All methods of order q share the important property that the local error $\varepsilon_k(t)$ of the solution is proportional to h^{q+1}.
- This property implies that changing the stepsize h by a factor of r changes the local error by a factor of r^{q+1}. It is this characteristic which makes high-order methods so efficient.

Table 7.5. Error function e_r for multistep and collocation methods of order q for equations of order $n = 2$

	Multistep				Collocation	
	Extrapolation		Interpolation			
q	$e_r(1)$	$\frac{de_r}{d\tau}(1)$	$e_r(1)$	$\frac{de_r}{d\tau}(1)$	e_r(q-n)	$\frac{de_r}{d\tau}(q-n)$
2	0.1667	0.5000	−0.3333	−0.5000	0.1667	0.5000
3	0.1250	0.4167	−0.0417	−0.0833	0.0000	0.3333
4	0.1056	0.3750	−0.0194	−0.0417	0.1500	0.3750
5	0.0937	0.3486	−0.0118	−0.0264	0.0000	0.3111
6	0.0856	0.3299	−0.0081	−0.0188	0.1364	0.3299
7	0.0796	0.3156	−0.0060	−0.0143	0.0000	0.2929
8	0.0749	0.3042	−0.0047	−0.0114	0.1263	0.3042
9	0.0710	0.2949	−0.0038	−0.0094	0.0000	0.2791
10	0.0679	0.2870	−0.0032	−0.0079	0.1185	0.2870
11	0.0652	0.2802	−0.0027	−0.0068	0.0000	0.2683
12	0.0628	0.2743	−0.0023	−0.0059	0.1123	0.2743
13	0.0608	0.2690	−0.0020	−0.0052	0.0000	0.2597
14	0.0590	0.2644	−0.0018	−0.0047	0.1072	0.2644

- From all methods included in Table 7.5 the multistep algorithms with one or more iterative improvement steps using eqns. (7.103) and matrix (7.106) clearly is the best. One may, however, expect that the pure extrapolation and the interpolation procedures are of comparable accuracy if the ratio of the stepsizes is
$$\frac{h_{\text{extr}}}{h_{\text{inter}}} = \left(\frac{0.2644}{0.0047}\right)^{1/(q+1)} \approx 1.3 \ . \tag{7.190}$$
- As the multistep method with one iteration step needs twice the number of evaluations of the right-hand sides of the differential equation w.r.t. the pure extrapolation method, the choice of the interpolation method is not obvious if only the approximation errors are considered. Multistep methods with interpolation are, however, the right choice if the accumulation of rounding errors is critical (i.e., for integrations involving billions of steps), these methods allow for longer stepsizes and therefore minimize the approximation errors.
- It is interesting to note that $e_r\,(q-n) = 0$ for odd orders of the collocation method. This fact might generate problems when making the attempt to develop error criteria, which *also* try to limit the error of the solution vector itself.
- The error behavior documented by Table 7.5 should be compared to the error of a Taylor series truncated after the terms of order q. Assuming that the terms of order $i \leq q$ could be computed “error-free” (as it can be done for simple differential equations), the local error-term associated with a Taylor series method simply reads as

$$\varepsilon_{r,\text{taylor}} = \frac{1}{(q+1)!}\, h_k^{q+1}\, \boldsymbol{y}^{(q+1)} . \tag{7.191}$$

Comparing this to the performance of the collocation method we obtain the ratio

$$\left| \frac{\varepsilon_{r,\text{colloc}}}{\varepsilon_{r,\text{taylor}}} \right| = e_{r,\text{colloc}}(q-n)\, \frac{(q+1)!}{(q-n)^{q+1}} \approx\ 2 \cdot 10^{-5}\ , \tag{7.192}$$

where the numerical value corresponds to $q = 12$. We obtain the surprising result that the error behavior of the collocation method is much better than that of the Taylor series solution of the same order. The comparison is about fair: In both methods $q+1$ terms (either the derivatives at t_k or the right-hand sides $\boldsymbol{f}(t_{k_j})$ of the differential equation systems) have to be calculated "from scratch".

- The same comparison could be made for the multistep method. The result reads as follows:

$$\left| \frac{\varepsilon_{r,\text{multi}}}{\varepsilon_{r,\text{taylor}}} \right| = e_{r,\text{multi}}(1)\ (q+1)! \approx 1.7 \cdot 10^{9}\ , \tag{7.193}$$

which of course clearly speaks in favor of the Taylor series method! The comparison is unfair, however: The multistep methods only need to evaluate the right-hand sides of the differential equation systems once per step (for the pure extrapolation method), whereas q derivatives have to be calculated in the case of the Taylor series.

The Error Function of the Two-body Problem. In Celestial Mechanics it is important to know the local error associated with the two-body solution. The circular orbit is an important and simple special case. Its derivatives are easily calculated as follows:

$$\begin{aligned} \boldsymbol{r}^{(2i)} &= (-1)^i\ n^{2i}\ \boldsymbol{r}\ , \quad i = 0,1,2,\ldots, \\ \boldsymbol{r}^{(2i+1)} &= (-1)^i\ n^{2i}\ \dot{\boldsymbol{r}}\ , \quad i = 0,1,2,\ldots\ . \end{aligned} \tag{7.194}$$

Using the approximation (7.189) for the local error function (of the multistep method) we obtain

$$\begin{aligned} \boldsymbol{\varepsilon}_k(t_k+h) &= e_r(1)\ (nh)^q\ h \begin{cases} (-1)^{q/2}\ \dot{\boldsymbol{r}}\ , & q \text{ even} \\ (-1)^{(q+1)/2}\ n\ \boldsymbol{r}\ , & q \text{ odd} \end{cases} , \\ \dot{\boldsymbol{\varepsilon}}_k(t_k+h) &= \frac{de_r(1)}{d\tau}\ (nh)^q \begin{cases} (-1)^{q/2}\ \dot{\boldsymbol{r}}\ , & q \text{ even} \\ (-1)^{(q+1)/2}\ n\ \boldsymbol{r}\ , & q \text{ odd} \end{cases} , \end{aligned} \tag{7.195}$$

where the subscript "multi" was left out.

This result allows it to study the error propagation when integrating a circular orbit. As usual, we are particularly interested in the impact of the local approximation error on the semi-major axis a. Adapting formula (7.173) to the approximation errors we obtain:

$$\delta a_k = \frac{2\,a^2}{\mu} \left\{ \frac{\mu}{r_k^3}\,\boldsymbol{r}(t_k)\cdot\boldsymbol{\varepsilon}_k(t_k+h) \;+\; \dot{\boldsymbol{r}}(t_k)\cdot\dot{\boldsymbol{\varepsilon}}_k(t_k+h) \right\} \,. \tag{7.196}$$

Taking into account that the scalar product $\boldsymbol{r}\cdot\dot{\boldsymbol{r}} = 0$ for a circular orbit, the error of the semi-major axis is obtained by introducing the above approximation for $\boldsymbol{\varepsilon}_k$ and its first derivative into formula (7.173):

$$\delta a_k = 2\,a\,(nh)^q \begin{cases} (-1)^{q/2}\ \frac{de_r(1)}{d\tau} & ; \quad q \text{ even} \\ (-1)^{(q+1)/2}\, nh\, e_r\ ; & \quad q \text{ odd} \end{cases} \,. \tag{7.197}$$

Equation (7.197) promises that methods of an odd order q have to be preferred to methods of an even order for two-body orbits with small eccentricities, because the error in a is bound by h^{q+1} rather than h^q. This result only holds for the semi-major axis a – not for the other orbital elements (first integrals).

Eventually, we are able to check whether the results in Figure 7.27 are in accordance with theory. For Jupiter, $a \approx 5.208$ and therefore $n \approx 1.45\cdot 10^{-3}$. As the stepsize was $h = 40$ days, the order $q = 14$, and as a multistep method with one interpolation step was used, we expect an error of

$$\delta a_k = 2\,a\,(nh)^q\,(-1)^{q/2}\,\frac{de_r(1)}{d\tau} \approx -2.39\cdot 10^{-19} \tag{7.198}$$

in the semi-major axis a per integration step. Observe, that in each step the same error is made. Therefore the expected accumulated error after one million years simply is the number of steps times the value in the above formula:

$$\delta a(t_N) = 2\,N\,a\,(nh)^q\,(-1)^{q/2}\,\frac{de_r(1)}{d\tau} \approx -2.18\cdot 10^{-12}\,, \tag{7.199}$$

where $N = \frac{1000000\cdot 365.25}{40} \approx 9131250$.

A negative drift (as observed in Figure 7.27) in the semi-major axis a is predicted by our theory. The order of magnitude, however, is disappointingly wrong: The error according to formula (7.199) is by about a factor of 50 too small to explain the actual errors in Figure 7.27.

This failure of theory is uniquely due to the small eccentricity of $e = 0.048$ of Jupiter's orbit! Repeating the same integration with an orbit of $e = 0$ would confirm the law (7.199). (Actually, one would no longer observe the accumulated approximation error, but the accumulated rounding error, when performing the integration with $h = 40$ and $e = 0$; the law may be established, e.g., with $h = 60$).

It is therefore important to note, that even for orbits with small eccentricities with $e < 0.1$ a useful stepsize cannot be predicted with a formula of type (7.199), which is based on the assumption of a circular orbit.

In order to obtain a useful formula we have to compute the mean values of the following scalar products over one revolution period U:

$$\chi_1'(q,e) \overset{\text{def}}{=} \frac{1}{U}\left|\int\limits_0^U \boldsymbol{r}(t)\boldsymbol{r}^{(q)}(t)\ dt\right| \quad ; \qquad \chi_2'(q,e) \overset{\text{def}}{=} \frac{1}{U}\left|\int\limits_0^U \dot{\boldsymbol{r}}(t)\boldsymbol{r}^{(q)}(t)\ dt\right| \ . \tag{7.200}$$

It turns out, that for even orders q only $\chi_1 \neq 0$, for odd orders only $\chi_2 \neq 0$. It is therefore possible to define the amplification factor

$$\chi(q,e) \overset{\text{def}}{=} \begin{cases} \chi_1'(q,e)/\chi_1'(q,0)\ ; & q \text{ odd} \\ \chi_2'(q,e)/\chi_2'(q,0)\ ; & q \text{ even} \end{cases} \ . \tag{7.201}$$

Averaged over a revolution, the change in the semi-major axis a per step may therefore simply be obtained from the corresponding relation (7.197) for circular orbits by multiplying it with the amplification factor $\chi(q+1,e)$:

$$\overline{\delta a_k} = 2\,a\,(nh)^q\,\chi(q+1,e) \begin{cases} (-1)^{q/2}\,\frac{de_r(1)}{d\tau} & ; \quad q \text{ even} \\ (-1)^{(q+1)/2}\,nh\,e_r(1)\ ; & \quad q \text{ odd} \end{cases} \ . \tag{7.202}$$

The amplification factors $\chi(q,e)$ are contained in Table 7.6. They are based on the computation of the Taylor series coefficients using the algorithm (7.27).

For methods with constant stepsize, the accumulated error in the semi-major axis a may now be calculated from the mean error (7.202) per step simply by multiplying this equation with the number N of integration steps in the entire interval:

$$\delta a(T_N) = 2\,a\,(nh)^q\,\chi(q+1,e)\,N \begin{cases} (-1)^{q/2}\,\frac{de_r}{d\tau} & ; \quad q \text{ even} \\ (-1)^{(q+1)/2}\,nh\,e_r\ ; & \quad q \text{ odd} \end{cases} \ . \tag{7.203}$$

If we apply this result to the integration underlying Figure 7.27 we obtain ($h = 40$, $q = 14$, $e = 0.048 \approx 0.05$, multistep-method with interpolation):

$$\delta a(t_N) = 2\,a\,(nh)^q\,\chi(q{+}1,e)\,N\,\frac{de_r}{d\tau} = -2.39{\cdot}10^{-19}{\cdot}87.525\,N \approx -190{\cdot}10^{-12}\ , \tag{7.204}$$

giving eventually the correct order of magnitude for the accumulated error actually observed in Figure 7.27 (!).

Using the relations $N = \frac{t_N}{h}$, $\delta n(t_N) = -\frac{3}{2}\,\frac{n}{a}\,\delta a(t_N)$ it is now easy to compute the error in the mean anomaly $\sigma(t)$ as an integral of $\delta n(t)$ over time t:

Table 7.6. Amplification of absolute values of derivatives of an orbit with eccentricity $e = 0.05$ w.r.t. derivatives of a circular orbit

	Mean Amplification of Errors $\chi(e)$ in a over Revolution					
q	$\chi(0)$	$\chi(0.025)$	$\chi(0.05)$	$\chi(0.075)$	$\chi(0.100)$	$\chi(0.125)$
1	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.002	1.008	1.017	1.031	1.048
4	1.000	1.002	1.008	1.017	1.031	1.048
5	1.000	1.009	1.038	1.087	1.158	1.255
6	1.000	1.009	1.038	1.087	1.158	1.255
7	1.000	1.040	1.163	1.382	1.718	2.209
8	1.000	1.040	1.163	1.382	1.718	2.209
9	1.000	1.162	1.689	2.704	4.450	7.351
10	1.000	1.162	1.689	2.704	4.451	7.357
11	1.000	1.668	4.043	9.399	20.504	42.898
12	1.000	1.668	4.043	9.400	20.523	43.053
13	1.000	3.827	15.871	50.818	143.741	379.445
14	1.000	3.827	15.872	50.855	144.292	384.361
15	1.000	13.710	87.525	383.631	1418.391	4753.541
16	1.000	13.710	87.549	384.682	1436.464	4934.982
17	1.000	65.100	638.604	3888.377	18871.046	80547.853
18	1.000	65.101	639.248	3922.941	19543.482	88014.411

$$\delta\sigma(t) = -\frac{3}{2}\,\frac{n}{a}\,\frac{1}{h}\int\limits_{t_0}^{t} \delta a(t')\; t'\; dt' \;, \tag{7.205}$$

where $\delta a(t)$ has to be replaced by eqn. (7.202). Formula (7.205) actually explains Figure 7.27 (bottom). (For small eccentricities we may consider the mean anomaly $\sigma(t)$ and the argument of latitude $u(t)$ as equivalent.)

The local approximation error was given above for the multistep method. For the collocation method h has to be replaced by $\frac{h_k}{q-n}$ and $e_r(1)$ by $e_r(q-n)$.

7.7.4 A Rule of Thumb for Integrating Orbits of Small Eccentricities with Constant Stepsize Methods

Based on our knowledge of the accumulation of rounding and approximation errors, and based on our analysis of the impact of the eccentricity, we are now able to derive a formula for the optimum stepsize $h_{\rm opt}$ promising the most accurate results in the integration interval of length Δt for all methods based on the collocation principle. We consider the stepsize h as optimal if, at the end of the integration interval, the accumulated rounding error and the accumulated approximation error in the semi-major axis are of the same order of magnitude.

We have seen that the propagation of rounding errors (root of variance) in a obeys a law of type

$$\delta a(T_N) = c_r \sqrt{N} \ , \tag{7.206}$$

whereas the propagation of approximation errors obeys a law of type

$$\delta a(T_N) = c_a \, N \, (nh)^q \ , \tag{7.207}$$

where, for the sake of simplicity, we assumed that the order q is even. c_a and c_r are known constants. Assuming constant stepsize (which makes sense for orbits with small eccentricities) we may replace N by $N = \frac{\Delta t}{h}$, which leads to the following formula for the optimum stepsize

$$n \, h_{\text{opt}} = \left(\frac{c_r}{c_a \sqrt{n \Delta t}} \right)^{\frac{2}{2q-1}} \ . \tag{7.208}$$

This equation may be brought into a form which is very easy to interpret by replacing the stepsize by the optimum number of integration steps per revolution

$$N_{\text{st/rev}} = \frac{2\pi}{n \, h_{\text{opt}}} \ , \tag{7.209}$$

and by using on the right-hand side the approximation

$$n \, \Delta t = 2\pi \, N_{\text{rev}} \ , \tag{7.210}$$

where N_{rev} is the total number of revolutions. Formula (7.208) may thus be replaced by

$$N_{\text{st/rev}} = 2\pi \left(\frac{c_a \sqrt{2\pi \, N_{\text{rev}}}}{c_r} \right)^{\frac{2}{2q-1}} \ . \tag{7.211}$$

According to eqn. (7.180) we have

$$c_r = \sqrt{\tfrac{2}{3}} \, \varepsilon_m \, b^{e_{0,\max}} \approx 7.25 \cdot 10^{-16} \ , \tag{7.212}$$

whereas eqn. (7.203) provides

$$c_a = 2 \, a \, \chi(q+1, e) \left| \frac{de_r(1)}{d\tau} \right| \approx 4.28 \ . \tag{7.213}$$

The numerical values are given for the example of Figure 7.27.

For the integration interval of one million years, the number of revolutions for Jupiter is $N_{\text{rev}} \approx \frac{1000000}{11.86} \approx 84317$. As the semi-major axis is $a \approx 5.208$ AU, formula (7.211) gives the result

$$N_{\text{st/rev}} = 2\pi \left(\frac{c_a \sqrt{2\pi \, N_{\text{rev}}}}{c_r} \right)^{\frac{2}{2q-1}} \approx 151 \ , \tag{7.214}$$

corresponding to a stepsize of about 29 days for multistep procedures with (at least) one interpolation step.

As we can see the tests were performed with "reasonable" stepsizes. The rule of thumb (7.211) may be used to answer a few questions:

- For multistep methods without interpolation the stepsize should be, as established above, about a factor of 1.3 shorter.
- For collocation methods the stepsize is a factor of $q - n$ longer.
- What would be the optimum stepsize, if the integration interval would be stretched by a factor of 1000? (integrating over one billion (10^9) years instead of one million years in our example)?

  The change is not dramatic: The only entry changing in formula (7.211) is the number of revolutions N_{rev} , which is a linear function of time. All in all, the stepsize has to be reduced by a factor of only $1000^{\frac{1}{2q-1}} \approx 1.3$, corresponding to $h \approx 22$ days in our example.
- How does the optimum stepsize depend on the eccentricity e of the orbit?

  If the eccentricity would be $e = 0.125$ instead of 0.05, Table 7.6 tells that the nominator of the rule of thumb (7.211) would be increased by a factor of about 54, implying that the stepsize would have to be reduced by a factor of about 1.35 .

  If we would have integrated a precisely circular orbit the nominator in formula (7.211) would be decreased by a factor of 87.5 , implying that the stepsize could be increased by a factor of 1.4 .

  Compared to the circular orbit, the stepsize for the integration of an elliptical orbit with $e = 0.125$ has to be roughly cut in half compared to the optimum stepsize for a circular orbit.

Note that the rule of thumb (7.211) does not say anything about the size of the accumulated errors. For this purpose we have to use the rules for the accumulation of rounding errors. As these are of the order of a few 10^{-12} after one million years in the semimajor axis, we have to expect an increase in the error level of the semi-major axis by a factor of $\sqrt{1000 \cdot 1.3} \approx 36$, one of $\sqrt{(1000 \cdot 1.3)^3} \approx 4.7 \cdot 10^4$ in the mean anomaly (and, in view of the small value for the eccentricity also in the argument of latitude) when integrating over one billion years.

7.7.5 The General Law of Error Propagation

So far, we followed in essence the principles published by Brouwer [26] in 1937 to study error propagation. The error accumulation in the first integrals was studied first, then the error approximation in the argument of latitude

u could be derived as a consequence of the error in the semi-major axis a. Following this procedure, useful formulae could be established for both, the accumulated rounding errors and the accumulated approximation errors (in the argument of latitude u, the coordinates and velocities, and other derived quantities). The results are strictly valid for the two-body problem, they hold approximately for the perturbed motion *provided* the integration period is not of excessive length and *provided* the perturbations are small compared to the main term.

One must of course ask the question what happens with more general problems. The answer is in principle very simple: For both error sources the study of error propagation must be based on the *fundamental law of error propagation* (5.98) derived in Chapter 5. Before doing that it is appropriate to make a few general comments:

- *Approximation errors*: The difficulties involved in calculating the $(q+i)$-th derivatives, $i = 1, 2, \ldots,\ q \geq 10$, were considerable already in the case of the two-body problem. It is close to impossible to obtain sound estimates for these quantities for more general, non-trivial problems. On the positive side we may note that a general reduction of all stepsizes involved reduces these errors dramatically – the accumulated error is bound by a power law of the type h^{q+1} for a method of order q with constant stepsize (see also subsequent mathematical developments). It is therefore always possible to "eliminate" this error type – even when integrating over long time periods.
- *Rounding errors*: If some first integrals are algebraically known for a general dynamical problem, the error propagation in these integrals follows exactly the rules established in sections 7.7.3 and 7.7.4 for the osculating Keplerian elements. Figures II- 4.5 in Chapter II- 4 may serve as an example: The total energy E and the three components of the angular momentum $\boldsymbol{h}$ are known to be conserved quantities (see eqns. (3.47) and (3.40) in Chapter 3). The accumulated rounding errors in these quantities therefore must follow the $\sqrt{N}$-law, exactly as in the case of the two-body problem. Unfortunately it is not possible to derive a general law giving, e.g., the ecliptical longitudes of all planets as a function of the conserved energy (otherwise we would in essence have solved the N-body problem "analytically"). Therefore, the error accumulation in the total energy E and in the angular momentum vector $\boldsymbol{h}$ may be used only to check a posteriori whether the integration error is governed by rounding errors (where the $\sqrt{N}$-law must hold) or whether the approximation errors still played a role (in which case a systematic pattern would be observed). Tests of this kind may be used to establish the optimal processing strategy in a given computer environment. They may *not* be used to assess the error in the arguments of latitude, coordinates and velocities, etc. of the planets over long time periods (typically longer than a few thousand years).

For general problems in dynamics and for advanced problems in Celestial Mechanics (integration over long time periods, general N-body problem, etc.) the study of error propagation has to be based on the *fundamental law of error propagation* (5.98):

$$\varDelta\tilde{\mathbf{Z}}(t_N) \overset{\text{def}}{=} \sum_{k=0}^{N} \varDelta\tilde{\mathbf{Z}}_k(t_N) = \tilde{\mathbf{Z}}(t_N) \sum_{k=0}^{N} \tilde{\mathbf{Z}}^{-1}(t_k)\ \tilde{\boldsymbol{\varepsilon}}_k \ .$$

Equation (5.98) represents the accumulated error $\varDelta\tilde{\mathbf{Z}}(t)$ of the state vector $\boldsymbol{y}(t)$ of a dynamical system at an epoch t_N as a linear combination of the local errors $\tilde{\boldsymbol{\varepsilon}}_k$ introduced at each boundary epoch of the sub-intervals I_k of the total integration interval $I = [t_0, t_N]$. These errors may be either approximation or rounding errors. The elements of the matrix $\tilde{\mathbf{Z}}(t_N)$ are the components of the solution vector (or its $n-1$ derivatives) at t_N of the $n\,d$ homogeneous variational equations associated with the initial values of the dynamical problem considered. Let us briefly sketch in the concluding two sections how to use the above equation to deal with the propagation of approximation and rounding errors.

Propagation of Approximation Errors: The General Law. Leaving out all error terms of higher than $(q+2-n)$-th order in the stepsize h_k , the approximation error at epoch t_k may be written as:

$$\tilde{\boldsymbol{\varepsilon}}_k = h_k^{q+2-n} \begin{pmatrix} \mathbf{0} \\ \cdots \\ \cdots \\ \mathbf{0} \\ \frac{d^{(n-1)}}{d\tau^{(n-1)}}\,(e_{r_k})\ \boldsymbol{y}_{k0}^{(q+1)} \end{pmatrix} \overset{\text{def}}{=} h_k^{q+2-n}\,\tilde{\boldsymbol{\varepsilon}}_{rk} \ , \tag{7.215}$$

where $\boldsymbol{\varepsilon}_{rk}^{(i)} \overset{\text{def}}{=} \frac{\boldsymbol{\varepsilon}_k^{(i)}}{h^{q+1-i}}$ is the normalized local approximation error, which is proportional to the $(q+1)$-st derivative of the solution vector $\boldsymbol{y}_k(t)$ and to a factor depending on the solution method (but not on the stepsize).

Assuming constant stepsize, we easily obtain the accumulated approximation error by introducing the above error term (7.215) into the general formula (5.98):

$$\begin{aligned} \varDelta\mathbf{Z}(t_N) &= h^{q+2-n}\ \tilde{\mathbf{Z}}(t) \sum_{k=0}^{N} \tilde{\mathbf{Z}}^{-1}(t_k)\ \tilde{\boldsymbol{\varepsilon}}_k \\ &= h^{q+1-n}\ \tilde{\mathbf{Z}}(t) \int_{t_0}^{t_N} \tilde{\mathbf{Z}}^{-1}(t')\ \tilde{\boldsymbol{\varepsilon}}(t')\ dt' \ , \end{aligned} \tag{7.216}$$

where the sum on the right-hand side of eqn. (7.216) was approximated by an integral (assuming that the approximation errors at the concrete epochs t_k may be interpreted as values of a function $\tilde{\boldsymbol{\varepsilon}}(t)$).

If an integration procedure with automatic stepsize control is used, formula (7.216) has to be modified slightly. Let us assume that the stepsize at t_k may be expressed by

$$h_k \overset{\text{def}}{=} s_k \, h \ , \tag{7.217}$$

where h is a meaningful reference stepsize, e.g., the mean stepsize for quasi-periodic problems.

This leads to the following modification of formula (7.216):

$$\begin{aligned} \Delta\mathbf{Z}(t_N) &= h^{q+2-n} \ \tilde{\mathbf{Z}}(t) \sum_{k=0}^{N} s_k^{q+2-n} \ \tilde{\mathbf{Z}}^{-1}(t_k) \ \tilde{\boldsymbol{\varepsilon}}_k \\ &= h^{q+1-n} \ \tilde{\mathbf{Z}}(t) \int\limits_{t_0}^{t_N} s^{q+1-n}(t') \ \tilde{\mathbf{Z}}^{-1}(t') \ \tilde{\boldsymbol{\varepsilon}}(t') \ dt' \ . \end{aligned} \tag{7.218}$$

Formulae (7.216) and (7.218) show that the accumulated error of an integration procedure of order q is proportional to h^{q+1-n} despite the fact that the local errors are of the order of $q+2-n$ in h. They also show that the errors of all derivatives $\boldsymbol{y}^{(i)}$, $i = 0, 1, \ldots, n-1$, are of the same order $q+1-n$ (and not, as one might believe, of the order $q+1-i$ for derivative i).

It must be pointed out that the computation of the accumulated error requires the knowledge of (or a fair estimate of) the local error function $\boldsymbol{\varepsilon}_k$, which in term depends of the $(q+1)$st derivative of the solution function. This is why the above formulae are of greater theoretical than practical importance.

Propagation of Rounding Errors: The General Law. Equation (5.98) is the fundamental equation for studying the accumulation of rounding errors as well. As opposed to the accumulation of approximation errors we have to consider the local rounding errors $\tilde{\boldsymbol{\varepsilon}}_k$ as values of random variables. In other words, we have to form the expectation value and the variance of the fundamental equation (5.98), assuming the corresponding quantities of the local rounding errors are known. The statistical properties of the local rounding errors have been defined by eqns. (7.171). Reminding ourselves of the notations (5.94) we may rewrite eqns. (7.171) as

$$E(\tilde{\boldsymbol{\varepsilon}}_k) = \mathbf{0} \tag{7.219}$$

and

$$\operatorname{var}(\tilde{\boldsymbol{\varepsilon}}_k) = \mathbf{D}_k \ , \tag{7.220}$$

where matrix $\mathbf{D}_k$ is a diagonal matrix, the diagonal elements of which are given by the diagonals of the matrices $\operatorname{var}(\tilde{\boldsymbol{\varepsilon}}_k)$ in eqns. (7.171).

We are now in a position to calculate the expectation value of equation (5.98):

$$E\big(\mathit{\Delta}\mathbf{Z}(t_N)\big) = \tilde{\mathbf{Z}}(t) \sum_{k=0}^{N} \tilde{\mathbf{Z}}^{-1}(t_k)\ E\left(\tilde{\varepsilon}_k\right) = \mathbf{0}\ , \tag{7.221}$$

i.e., the expectation value for the accumulated rounding error is zero, as well.

The variance of the accumulated rounding error is obtained as the expectation value of the Cartesian product $\mathit{\Delta}\mathbf{Z}(t) \otimes \mathit{\Delta}\mathbf{Z}^T(t)$:

$$\begin{aligned}
\mathrm{var}\big(\mathit{\Delta}\mathbf{Z}(t_N)\big) &= E\Big(\mathit{\Delta}\mathbf{Z}(t_N) \otimes \mathit{\Delta}\mathbf{Z}^T(t_N)\Big) \\
&= \tilde{\mathbf{Z}}(t_N) \left[\sum_{k=0}^{N} \tilde{\mathbf{Z}}^{-1}(t_k)\ \mathrm{var}\left(\tilde{\varepsilon}_k\right) \big(\tilde{\mathbf{Z}}^{-1}(t_k)\big)^T\right] \tilde{\mathbf{Z}}^T(t_N) \\
&= h^{-1}\, \tilde{\mathbf{Z}}(t_N) \\
&\quad \cdot \left[\int_{t_0}^{t_N} s^{-1}(t')\ \tilde{\mathbf{Z}}_0^{-1}(t')\ \mathrm{var}\big(\tilde{\varepsilon}(t')\big)\big(\tilde{\mathbf{Z}}^{-1}(t')\big)^T\ dt'\right] \tilde{\mathbf{Z}}^T(t_N)\ ,
\end{aligned} \tag{7.222}$$

where $h_k = s_k\, h$ and $h(t) = s(t)\, h$ are defined as in the previous section, when the accumulation of approximation errors was studied.

Equation (7.222) requires the knowledge of the local variances $\mathrm{var}\big(\boldsymbol{\varepsilon}(t')\big)$. As opposed to the case of approximation errors it is possible to come up with fair estimates for these quantities during the integration process. If the variational equations are solved together with the primary equations, formula (7.222) might very well be used to compute the variance-covariance matrix of the accumulated rounding error. It is also possible to compute only the diagonal terms, the roots of which are representative for the errors of the components of the solution vector and its derivatives.

Equation (7.222) also shows, that as soon as a problem is governed by rounding errors (and no longer by approximation errors) the accumulated error must grow *at least* according to $(\sqrt{h})^{-1}$ or $\sqrt{N}$. The actual error propagation law is of course governed by the entire expression (7.222).

When comparing the solutions of the variational equations associated with a perturbed trajectory with those of the corresponding two-body orbit (compare, e.g., Figures 5.1 and 5.2) the limitations of using the two-body approximation (instead of the formulae developed in this final section) for studying the error propagation become apparent: Order of magnitude differences between the two types of solutions may easily occur. They are the rule when studying resonance phenomena (see, e.g., section II- 4.3.4).

8. Orbit Determination and Parameter Estimation

8.1 Orbit Determination as a Parameter Estimation Problem

Orbit determination must be viewed a special case of a general parameter estimation problem, where the *parameters* characterizing the orbit of a minor planet or of an artificial Earth satellite, have to be determined from *observations* of these celestial bodies.

Observations are – apart from the unavoidable observation errors – values of functions (the so-called *observed functions*) of the parameter estimation problem considered. In our application the observed functions are *nonlinear* in the orbit parameters. *Nonlinearity* and the fact that initially there may be *no approximate values available for the orbit parameters* are the essential difficulties of the orbit determination problem.

It is assumed that the readers of this book are familiar with the principles of parameter estimation theory, in particular with the basic facts of the method of least squares, which is the basis for the subsequent treatment. These basic facts are briefly reviewed, so to speak *en passant*, using the orbit improvement problem as an example in section 8.2 and when introducing the general parameter estimation problem of satellite geodesy in section 8.5.

The orbit parameters must uniquely specify one particular solution of the equations of motion and (possibly) the force field acting on the object. In a *pure orbit determination problem* the forces acting on the bodies are assumed to be known as a function of the bodies' positions (and possibly velocities). In this case, the orbit parameters are uniquely quantities defining the initial state (position- and velocity-vector of the orbit), at a particular epoch t_0. Whether or not a problem may be described by a *pure* orbit determination problem heavily depends on the time interval covered by the observations. It is, e.g., clear that many *dynamical parameters* (definition in section 5.2) have to be set up, if a high-accuracy orbit of a LEO (cm-dm accuracy) has to be established over one day (with about $15-20$ revolutions) using the observations of a spaceborne GPS receiver. No dynamical parameters have to be estimated if the observations within one opposition of a minor planet are analyzed. If initial values and dynamical parameters have to be determined, we speak of a *general orbit determination problem.* The most general parameter estimation problem considered in this chapter does not only contain orbit parameters, but also parameters related to the observers' (observatories') trajectories in inertial space. Such observer-specific parameters are, e.g., the coordinates of the observers in an Earth-fixed reference system, the Earth rotation parameters, parameters characterizing the Earth's atmosphere, etc.

Orbit determination and parameter estimation in Celestial Mechanics is an extremely broad field, which would yield enough material to fill an entire textbook. In this chapter we can therefore only address the most important aspects related to this topic. Section 8.2 introduces the problems, sections 8.3 and 8.4 deal with classical orbit determination and orbit improvement, where the attribute *classical* stands for the use of astrometric positions as observations and for using an extremely simple (and a priori known) force model, often even that of the two-body motion. In section 8.5 the scope is broadened. The general parameter estimation task of satellite geodesy is defined and a short overview of the particularities of the analysis in space geodesy, in particular of analyzing SLR/LLR and GPS data, is provided. The chapter concludes with a relatively new type of orbit determination problem, namely that of deriving LEO orbits using the measurements of onboard GPS receivers. Pure kinematic orbits (neglecting the equation of motion), pure dynamic orbits (assuming that the force field acting on the satellite is perfectly known) and mixtures of the two orbit types are considered.

8.2 The Classical Pure Orbit Determination Problem

In the planetary system classical orbit determination may be defined as the

> Estimation of the (osculating) orbital elements $a\,,e\,,i\,,\Omega\,,\omega\,,$ and T_0 of a celestial body referring to a particular epoch t_0 from a series of topocentric observations $t_i\,;\alpha'_i\,,\delta'_i\,,\ i=1,2,\ldots,n\geq 3$. (8.1)

The orbital elements a, e, i, Ω, ω, and T_0 were defined in section 4.1. The definition (8.1) implies that orbit determination is the inverse problem of the *computation of ephemerides*, where topocentric or geocentric spherical coordinates $\alpha_i\,,\delta_i\,,\Delta_i\,,\ i=1,2,\ldots,n\,,$ of a celestial body as observable at epochs t_i are computed from a given set of osculating elements (see Chapter 4). α_i and δ_i are the right ascensions and declinations (to be defined more precisely below) at the tabular epochs t_i ; the Δ_i are (in essence) the body's topocentric distances at the epochs t_i .

Observations from one or more observatories at different locations on the Earth (and/or elsewhere) may be used for orbit determination. It only matters that the heliocentric position vectors of the observer(s) at the observation times are known.

The osculating orbital elements in definition (8.1) should be understood as parameters defining the initial values at time t_0 in a problem governed in the most general case by equations of motion of type (3.21) or (3.143), in the simplest case by the equations of motion (4.1) of the two-body problem. In the latter case the time-span covered by observations must be sufficiently short (in practice shorter than about half a revolution), because otherwise the perturbations would be visible as systematic effects in the residuals.

Orbit determination and ephemeris calculation are the two fundamental tasks in applied Celestial Mechanics. The former task is much more demanding than the latter one. Algorithms for the computation of ephemerides are simple and transparent: In the first step, a table of helio- or geocentric positions has to be compiled, in a second step the geo- or topocentric ephemerides are obtained by applying a series of geometric transformations (translations and rotations). Orbit determination, on the other hand, includes a non-linear parameter estimation process.

The orbit determination problem may be solved with standard procedures of applied mathematics, *provided* a set of approximate orbit parameters of a sufficiently high accuracy is already available. If such approximations are available we speak of an *orbit improvement problem*, because we "merely" have to improve the known approximative parameters. This task is considered in the following paragraph.

8.2.1 Solution of the Classical Orbit Improvement Problem

Assuming that

$$a^K, e^K, i^K, \Omega^K, \omega^K, \text{ and } T_0^K \tag{8.2}$$

is a known set of approximate values of the orbit parameters a, e, i, Ω, ω, and T_0, we may develop the observed functions $\alpha(t; a, e, i, \Omega, \omega, T_0)$ and $\delta(t; a, e, i, \Omega, \omega, T_0)$ into Taylor series about the values (8.2) as origin:

$$\begin{aligned}
\alpha(t; a, e, i, \Omega, \omega, T_0) &= \alpha^K(t) + \sum_{j=1}^{6} \left(\frac{\partial \alpha^K}{\partial I_j} \right)(t) \, \left(I_j - I_j^K \right) + O(I_k I_l) \\
\delta(t; a, e, i, \Omega, \omega, T_0) &= \delta^K(t) + \sum_{j=1}^{6} \left(\frac{\partial \delta^K}{\partial I_j} \right)(t) \, \left(I_j - I_j^K \right) + O(I_k I_l) \; .
\end{aligned} \tag{8.3}$$

The notation

$$\{I_1, I_2, \ldots, I_6\} \overset{\text{def}}{=} \{a, e, i, \Omega, \omega, T_0\} \tag{8.4}$$

was used to characterize the elements, and the notation

$$\begin{aligned}\alpha^K(t) &\overset{\text{def}}{=} \alpha\big(t\,; a^K, e^K, i^K, \Omega^K, \omega^K, T_0^K\big)\\ \delta^K(t) &\overset{\text{def}}{=} \delta\big(t\,; a^K, e^K, i^K, \Omega^K, \omega^K, T_0^K\big)\end{aligned} \tag{8.5}$$

to denote the right ascensions and declinations computed with the approximate orbit parameters (8.2).

The functions $\alpha^K(t)$, $\delta^K(t)$ and their partial derivatives w.r.t. the orbit parameters are computed from the solutions of the initial value problem associated with the primary equation (3.21), (3.143) or (4.1) and of the initial value problems associated with the variational equations corresponding to the orbit parameters (see Chapter 5).

Neglecting terms of second and higher order in eqns. (8.3) and subtracting from these equations the observations α_i' and δ_i' specified in the definition (8.1), we obtain the set of linear *observation equations*

$$\begin{aligned}\sum_{j=1}^{6} \frac{\partial \alpha_i^K}{\partial I_j}\big(I_j - I_j^K\big) - \big(\alpha_i' - \alpha^K(t_i)\big) \qquad &= v_{\alpha_i}\\ \sum_{j=1}^{6} \frac{\partial \delta_i^K}{\partial I_j}\big(I_j - I_j^K\big) - \big(\delta_i' - \delta^K(t_i)\big) \qquad &= v_{\delta_i}\\ i = 1, 2, \ldots, n\ ,&\end{aligned} \tag{8.6}$$

where

$$\frac{\partial \alpha_i^K}{\partial I_j} \overset{\text{def}}{=} \left(\frac{\partial \alpha^K}{\partial I_j}\right)(t_i) \quad \text{and} \quad \frac{\partial \delta_i^K}{\partial I_j} \overset{\text{def}}{=} \left(\frac{\partial \delta^K}{\partial I_j}\right)(t_i)\ . \tag{8.7}$$

The right-hand sides v_{α_i} and v_{δ_i} of eqns. (8.6) are called the *residuals* in right ascension and declination. Ideally, if the observations were error-free, these quantities would be equal to zero. In practice, the system of observation equations (8.6) is solved to make the sum of residuals squares a minimum:

$$\sum_{i=1}^{n} \left\{\big[\cos\delta_i'\, v_{\alpha_i}\,\big]^2 + v_{\delta_i}^2\right\} = \text{ min.}\ . \tag{8.8}$$

This particular minimum principle (8.8) is appropriate, if the observed angles $\cos\delta_i'\ \alpha_i'$ and δ_i' at t_i are of comparable accuracy, and if the accuracies are comparable for all observations $i = 1, 2, \ldots, n$. These assumptions are very closely met, if one and the same telescope is used to make photographic or CCD observations. If observations stemming from telescopes of different optical qualities (e.g., of different focal lengths) are analyzed together, the residuals must be *weighted* to take quality differences into account. The resulting minimum principle will, however, again minimize a sum of weighted residuals squares – only the coefficients (weights) of the terms are different from the simplest form (8.8):

$$\sum_{i=1}^{n} \frac{\sigma_0^2}{\sigma_i^2} \left\{ \left[\cos\delta_i' \, v_{\alpha,i} \right]^2 + v_{\delta,i}^2 \right\} = \text{ min. } , \tag{8.9}$$

where σ_i is the estimated rms error of observation number i, and σ_0 the so-called mean error of unit weight. The weight unit may in principle be chosen arbitrarily. In practice it makes sense to associate it with the mean observation error for one particular telescope. In order not to complicate the discussion, we will always use the simpler minimum principle (8.8), but keep in mind that generalizations of type (8.9) are possible without problems.

As the residuals in the expression (8.8) may be replaced by eqns. (8.6), the sum in eqn. (8.8) is actually a quadratic form in the orbit parameters $\mathit{\Delta} I_j^K \overset{\text{def}}{=} I_j - I_j^K$, $j = 1, 2, \ldots, 6$. This quadratic form assumes a minimum, if its partial derivatives w.r.t. all six orbit parameters I_j are zero. The resulting system of six algebraic equations is linear. It is the well-known *system of normal equations*:

$$\mathbf{N}^K \, \mathit{\Delta}\boldsymbol{I}^K = \boldsymbol{b}^K \ , \tag{8.10}$$

where

$$\left(\mathit{\Delta}\boldsymbol{I}^K\right)^T \overset{\text{def}}{=} \left(a - a^K, e - e^K, i - i^K, \mathit{\Omega} - \mathit{\Omega}^K, \omega - \omega^K, T_0 - T_0^K\right) \ , \tag{8.11}$$

and where $\mathbf{N}^K$ is a quadratic and symmetric matrix, the general element of which is defined by

$$N_{jk}^K = \sum_{i=1}^{n} \left\{ \cos^2\delta_i' \, \frac{\partial\alpha_i^K}{\partial I_j} \, \frac{\partial\alpha_i^K}{\partial I_k} + \frac{\partial\delta_i^K}{\partial I_j} \, \frac{\partial\delta_i^K}{\partial I_k} \right\} \ ; \quad j, k = 1, 2, \ldots, 6 \ . \tag{8.12}$$

$\boldsymbol{b}^K$ is a column array of six elements, the general element of which is defined as

$$b_j^K = \sum_{i=1}^{n} \left\{ \cos^2\delta_i' \, \frac{\partial\alpha_i^K}{\partial I_j} \left(\alpha_i' - \alpha^K(t_i)\right) + \frac{\partial\delta_i^K}{\partial I_j} \left(\delta_i' - \delta^K(t_i)\right) \right\} , \quad j = 1, 2, \ldots, 6 \ . \tag{8.13}$$

The solution of the normal equation system (8.10) is usually written in the form

$$\mathit{\Delta}\boldsymbol{I}^K = \left(\mathbf{N}^K\right)^{-1} \boldsymbol{b}^K \overset{\text{def}}{=} \mathbf{Q}^K \, \boldsymbol{b}^K \ . \tag{8.14}$$

The variance-covariance matrix associated with the solution vector $\mathit{\Delta}\boldsymbol{I}$ is given by

$$\mathbf{cov}\left(\mathit{\Delta}\boldsymbol{I}^K\right) = \left(m_0^K\right)^2 \mathbf{Q}^K \ , \tag{8.15}$$

where the a posteriori variance factor $\left(m_0^K\right)^2$ of one observation is defined by

$$\left(m_0^K\right)^2 = \frac{\sum\limits_{i=1}^{n} \left\{ \left[\cos\delta_i' \, v_{\alpha,i}^K \right]^2 + \left(v_{\delta,i}^K\right)^2 \right\}}{2n - 6} \ . \tag{8.16}$$

The solution of each orbit determination task (8.1) is therefore accompanied by a full variance-covariance matrix, allowing it to compute the mean error of the determined orbit elements or of functions thereof, assuming that the number n of direction observations exceeds three.

So far, we have shown that the orbit determination (8.1) may be reduced to a standard least-squares procedure, provided that the non-linear parameter estimation problem may be described by the set (8.6) of linear equations in the unknowns. Whether or not this assumption holds, may be checked after having solved the normal equation system (8.10): One has to see whether the second and higher-order terms in the Taylor series expansion (8.3) actually could be neglected, i.e., whether they were small compared to the mean error m_0^K of the observations. In practice we have to check whether

$$\begin{aligned}
\left| \alpha(t_i; a, e, \ldots, T_0) \; - \; \left\{ \alpha^K(t_i) + \sum_{j=1}^{6} \frac{\partial \alpha_i^K}{\partial I_j} \left(I_j - I_j^K\right) \right\} \right| \cos^2 \delta_i' \; &\ll \; m_0^K \\
\left| \delta(t_i; a, e, \ldots, T_0) \; - \; \left\{ \delta^K(t_i) + \sum_{j=1}^{6} \frac{\partial \delta_i^K}{\partial I_j} \left(I_j - I_j^K\right) \right\} \right| \; &\ll \; m_0^K \; ,
\end{aligned} \tag{8.17}$$

where the terms $\alpha(t_i; a, e, \ldots, T_0)$ and $\delta(t_i; a, e, \ldots, T_0)$ have to be calculated using the values $a \overset{\text{def}}{=} a^K + \Delta a^K$, $e \overset{\text{def}}{=} e^K + \Delta e^K$, etc. for the orbital elements.

With the above developments we may define the solution of the orbit classical orbit improvement problem (8.1) through the following procedure:

Iterative Solution of the Classical Orbit Improvement Problem.

1. *Initialize* the process by defining/selecting a first set of approximate orbital elements, i.e.,

$$\text{for } K = 0 \text{ define } a^K, e^K, i^K, \Omega^K, \omega^K, T_0^K \; . \tag{8.18}$$

2. *Improve* the solution *iteratively* for $K = 1, 2, \ldots$ using the observation equations (8.6) and the normal equation system (8.10). Calculate the mean error m_0^K of the observations according to formula (8.16).
3. Calculate the terms on the left-hand sides of the expressions (8.17).
4. *Terminate* the orbit determination process by setting $I_j \overset{\text{def}}{=} I_j^K + \Delta I_j^K$, $j = 1, 2, \ldots, 6$, if all conditions (8.17) are met for $i = 1, 2, \ldots, n$.
5. If this is not the case, invoke a new iteration step by using the solutions $I_j^{K+1} \overset{\text{def}}{=} I_j = I_j^K + \Delta I_j^K$ as the initial values of the new iteration step $K + 1$ and proceed with step 2.

Orbit determination is thus nothing but a standard procedure of applied mathematics to treat *non-linear parameter estimation problems.* It is based on the *linearization* of the originally non-linear problem. The solution method is characterized by the keywords *initialization, iteration*, and *termination.*

The crucial part of non-linear parameter estimation procedures is the initialization step. If for $K = 0$ the available approximation is poor, the iteration process may diverge, or, if multiple solutions occur, the iterative solution may swap from one solution to (the) other(s). If a problem has more than one solution (i.e., if the sum (8.8) has several minima) and if the iteration process happens to converge, it may be rather questionable whether the correct solution was found. This is why the initialization of the orbit determination problem has to be studied carefully.

In the case of orbit determination the initialization step is called *first orbit determination* or *initial orbit determination. First* orbit determination is therefore nothing but the task to find an initial set of orbit parameters of sufficient quality for the above iterative orbit improvement process to converge. The result of a first orbit determination should not be viewed as some kind of a “final” result. The final result shall always be based on all observations available at the time of the analysis and it shall represent them in the sense of the method of least squares by applying a minimum principle of type (8.8) or (8.9).

The definition (8.1) of the orbit determination problem, when applied to the planetary system, may be modified in several ways: The equatorial coordinates (right ascension α and declination δ) might be replaced by ecliptical coordinates (ecliptical longitude λ and ecliptical latitude β), or by different (but equivalent) sets of orbital elements.

The definition (8.1) still lacks precision:

1. The observations involve the position and the velocity vectors of the observed object *and* of the observer(s). In the orbit determination problem, the position vectors of the observers (expressed in the coordinate system of the orbital elements) are assumed to be known.
2. The light propagates with velocity c in vacuum, which is why the observed angles α_i and δ_i only approximately define the unit vector at time t_i from the observer to the observed object. The geometric quantity derived from the observations is the so-called *astrometric position*, to be defined in paragraph 8.2.2.

With the understanding that clarifications are necessary and that modifications are possible, definition (8.1) still is a valid characterization of the *orbit determination problem in the planetary system.*

Classical orbit determination problems as defined by (8.1) are also encountered in satellite geodesy, e.g., in the context of optical surveys of space debris (see, e.g., [99], [101]), but so far it never reached the importance of the corresponding task in the planetary system. The reasons are manifold:

- Whereas direction observations play a dominant role for applications in the planetary system, other observation types like ranges or range rates (resulting, e.g., from the RADAR technique or from observing the Doppler-shifted signals emitted by satellites, or from the SLR technique) are of much greater importance in satellite geodesy.
- With each observation technique and each combination of such techniques one might formulate a "new" orbit determination problem. For applications of this kind the definition (8.1) should be generalized by allowing for more than one observation type. This is not done here, because these new problems never could attract similar attention as the classical orbit determination problem (8.1).
- For geodetic routine applications only the second part of the task, namely the orbit improvement part, is relevant. Usually, initial orbits of sufficient quality are available to invoke directly the orbit improvement process. (Often it is not even possible to observe the satellites without excellent orbit predictions).
- In scientific applications, orbit determination often has to be considered as a sub-task of a more general parameter estimation problem. The generalizations are manifold:
  - The orbit description always includes the initial osculating elements, but it may well be that *dynamical* parameters characterizing the force field (e.g., parameters of the Earth's gravity field, radiation pressure parameters, or empirical parameters) must be considered, as well.
  - In order to model the observations with sufficient accuracy, it may be necessary to determine not only orbit parameters, but also parameters defining the position of the observer (observatory coordinates) and the transformation parameters between the celestial and the terrestrial reference systems (Earth rotation and orientation parameters).
  - It may not be possible to formulate and solve one orbit determination problem per observed satellite, but only a general parameter estimation problem encompassing the orbit parameters for an entire system of satellites. This general approach is, e.g., required when analyzing the observations of navigation satellite systems like the US GPS or the Russian GLONASS.

Despite these remarks, which do somewhat attenuate the importance of first orbit determination in satellite geodesy, one should keep in mind that only direction observations allow it to determine the orbit of a celestial object using the observations from *only one* observatory.

In sections 8.3 and 8.4 the problem of *first orbit determination* and that of *orbit improvement* will be dealt with separately.

8.2.2 Astrometric Positions

The position of a celestial object P on the celestial sphere directly accessible to an observer O at observation time t (usually referred to as *apparent place*) is not easily interpreted geometrically, due to several reduction terms, namely

- the refraction of light in the Earth's atmosphere,
- the annual and diurnal aberration corrections due to the motion of the observer w.r.t. the inertial system,
- the gravitational deflection of light,
- the light propagation time $\frac{\Delta}{c}$ between P and O,
- etc.

It is, however, comparatively easy to derive the *astrometric position* of the celestial body at observation time t from the directly observed apparent place.

We define the astrometric position as the *geometric direction from the observer at time t to the celestial object at time $t - \frac{\Delta}{c}$*, where Δ is the distance between P at time $t-\frac{\Delta}{c}$ and O at time t. Figure 8.1 illustrates the astrometric position represented by the unit vector $\boldsymbol{e}(t)$. By using astrometric positions as illustrated in Figure 8.1 we avoid to include terms into the reduction process of photographic plates and CCD-arrays, which depend on the distance between the observed object and the observer. One should keep in mind that this distance usually is not known with sufficient precision when reducing the plates and/or arrays. When dealing with objects of the planetary system (like planets, minor planets, comets) the astrometric position is comparable to catalogue *mean places* of stars (see, e.g., glossary, term "astrometric ephemeris"

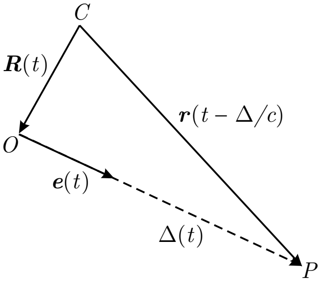


Fig. 8.1. Astrometric position $\boldsymbol{e}$ at observation time t of a celestial body P at a distance Δ from an observer O, and position vector $\boldsymbol{r}\left(t - \frac{\Delta}{c}\right)$ w.r.t. the center C of the primary attracting body

in the Explanatory Supplement [107]). When dealing with artificial Earth satellites (geocentric orbits), the astrometric position is in essence the sum of the mean place (represented by the catalogue mean places of stars) and the annual aberration.

The (astrometric) coordinates α and δ referring to the selected (quasi)-inertial coordinate system may be computed from the relations (see Figure 8.1)

$$\boldsymbol{\Delta}(t) = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \end{pmatrix} = \Delta \begin{pmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{pmatrix} \overset{\text{def}}{=} \boldsymbol{r}\left(t - \frac{\Delta}{c}\right) - \boldsymbol{R}(t) \ , \tag{8.19}$$

where we have assumed that the selected coordinate system is an equatorial system referring to a particular epoch (e.g., the system $J2000.0$). The astrometric right ascension and declination are obtained from the above relations as

$$\begin{aligned} \alpha(t) &= \arctan\left(\frac{\Delta_2}{\Delta_1}\right) \\ \delta(t) &= \arcsin\left(\frac{\Delta_3}{\Delta}\right) \ . \end{aligned} \tag{8.20}$$

In the orbit determination process we need the formulae (8.20) to compute the terms "observed - computed" in the observation equations. We also need the partial derivatives of these quantities w.r.t. the orbital elements. Due to the structure of the formulae (8.20) and due to the fact that the position vector $\boldsymbol{R}(t)$ of the observer does *not* depend on the orbit parameters, we obtain the following simple formulae for these partial derivatives:

$$\begin{aligned} \frac{\partial\alpha}{\partial I} &= \nabla_\Delta \alpha \cdot \frac{\partial \boldsymbol{r}\left(t - \frac{\Delta}{c}\right)}{\partial I} \\ \frac{\partial\delta}{\partial I} &= \nabla_\Delta \delta \cdot \frac{\partial \boldsymbol{r}\left(t - \frac{\Delta}{c}\right)}{\partial I} \ , \end{aligned} \tag{8.21}$$

where $I \in \{a, e, i, \Omega, \omega, T_0\}$, and where the gradient has to be taken w.r.t. the components Δ_i .

With formulae (8.20) it is a straightforward matter to compute the gradients of α and δ:

$$\nabla_\Delta \alpha = \frac{1}{\Delta_1^2 + \Delta_2^2} \begin{pmatrix} -\Delta_2 \\ \Delta_1 \\ 0 \end{pmatrix} \quad \text{and} \quad \nabla_\Delta \delta = \frac{1}{\Delta^2 \sqrt{\Delta_1^2 + \Delta_2^2}} \begin{pmatrix} -\Delta_1\,\Delta_3 \\ -\Delta_2\,\Delta_3 \\ \Delta_1^2 + \Delta_2^2 \end{pmatrix} . \tag{8.22}$$

In view of the fact that the terms $\frac{\Delta}{c}$ are small (a few seconds in the planetary system, fractions of seconds for artificial Earth satellites), it is safe to use the approximation

$$\boldsymbol{r}\left(t - \frac{\Delta}{c}\right) \overset{\text{def}}{=} \boldsymbol{r}(t) \; - \; \frac{\Delta}{c}\, \dot{\boldsymbol{r}}(t) \; . \tag{8.23}$$

Consequently, the partial derivatives of the position vector $\boldsymbol{r}\left(t - \frac{\Delta}{c}\right)$ w.r.t. the orbit parameters are obtained by

$$\begin{aligned}\left(\frac{\partial \boldsymbol{r}}{\partial I}\right)\left(t - \frac{\Delta}{c}\right) &= \left(\frac{\partial \boldsymbol{r}}{\partial I}\right)(t) \; - \; \frac{\Delta}{c}\left(\frac{\partial \dot{\boldsymbol{r}}}{\partial I}\right)(t) \; - \; \left(\nabla_{\Delta}(\Delta(t)) \cdot \left(\frac{\partial \boldsymbol{r}}{\partial I}\right)(t)\right)\frac{\dot{\boldsymbol{r}}}{c} \\ &\approx \left(\frac{\partial \boldsymbol{r}}{\partial I}\right)(t) \; . \end{aligned} \tag{8.24}$$

Note that $\nabla_{\Delta}\Delta(t)$ is the unit vector from the observer to the observed object at observation time t. The latter approximation in the previous equation neglects even the terms $O\left(\frac{\Delta}{c}\right)$ of first order in the light propagation time. This approximation for the partial derivatives is usually good enough as an approximation (the corresponding approximation is *not* allowed in the terms "observed-computed"). The partials of the position vector w.r.t. the orbital elements $I_j\,,\; j = 1, 2, \ldots, 6$ emerge as solutions of the corresponding variational equations (see Chapter 5).

Subsequently we will use the program ORBDET, as documented in Chapter II- 8 of Part III, to illustrate the problem of orbit determination. The program processes astrometric positions either of planets and comets or of Earth satellites and space debris.

Table 8.1 shows the observations of minor planet *Silentium*, an object that was discovered by Paul Wild at the Zimmerwald Observatory (Station No. 026) on October 18, 1977. Its name *Silentium* witnesses the discoverer's attempt to obey the IAU recommendation to reduce the sometimes lengthy explanations for the minor planets' names. That the attempt was successful is documented in the *Dictionary of Minor Planet Names* [102]. The dictionary is highly recommendable from the scientific, the cultural, and the entertainment point of view.

Table 8.1 contains observations of the minor planet at the time of discovery in 1977 and of 1993. The 1977 and 1993 observations were made during opposition periods. The observations will be used to illustrate the basic concepts of orbit determination. It is important to note that observations from more than one observatory may be used. Note, that two pre-discovery positions of the minor planet were found a posteriori on photographic plates made at the Nauchnyj Observatory on Crimea. The observations of minor planets and comets may also be defined in a slightly different format, namely the one used by the Minor Planet Center (MPC), (60 Garden St., Cambridge MA 02138 USA) of the International Astronomical Union (IAU).

Table 8.1. Observations of minor planet *Silentium*

```
NAME: Silentium         REF.: MPC 4393, 5199, 5202, 23004
                        EQUINOX: 1950.0
JJJJ MM DD.DDDDD        HH MM SS.SS   SDD MM SS.S      STA
1977  9 19.05254         1 57 57.05    13 56 03.2      095
1977  9 22.04377         1 56 39.72    13 59 55.0      095
1977 10 18.91701         1 33 44.13    13 24 10.9      026
1977 10 19.98750         1 32 36.96    13 20 45.2      026
1977 10 20.03750         1 32 33.76    13 20 37.1      026
1977 11  3.78160         1 18 42.66    12 31 09.8      026
1977 11  5.81667         1 17 09.84    12 24 55.5      026
1977 11  9.00208         1 14 59.98    12 15 55.4      026
1977 11  9.04514         1 14 58.47    12 15 49.8      026
1977 11 10.91441         1 13 52.19    12 11 08.3      026
1977 11 11.01111         1 13 48.79    12 10 55.2      026
1993  9 19.92049         0 54 39.97     6 48 22.0      026
1993  9 20.94375         0 53 48.30     6 46 40.0      026
1993 10 10.92431         0 34 37.76     5 53 50.8      026
1993 10 11.96181         0 33 39.08     5 50 44.6      026
```

As mentioned, the program ORBDET is also capable of processing astrometric positions from artificial satellites. The observations have to be provided in a different format, which will be specified in Chapter II- 8.

8.3 First Orbit Determination

First orbit determination is the initialization step associated with the iterative solution of the orbit determination problem (8.1) as outlined in paragraph 8.2.1. The task has an artistic element: Intuition, imagination, elegance, but also opinions (strong ones, at times), etc. play an essential role. This is probably why protagonists of certain procedures at times assume the role of art-critics.

Two eminent mathematicians and astronomers, Gauss and Laplace (in alphabetic order), are considered the pioneers of the problem of first orbit determination in the sense the problem is understood today. The two names stand (in the above order) for two different approaches, namely *first orbit determination as a boundary value problem* and *as an initial value problem.* The two approaches are briefly reviewed in section 8.3.6 despite the fact that some of the concepts are outdated due to the computational resources available today.

In 1809 Gauss gave a very concise description of the task. The first few lines of Gauss' 1809 treatment [44] are reproduced in Figure 8.2. Today, we would probably call this part the abstract of the article.

There are four essential elements in the introductory text:

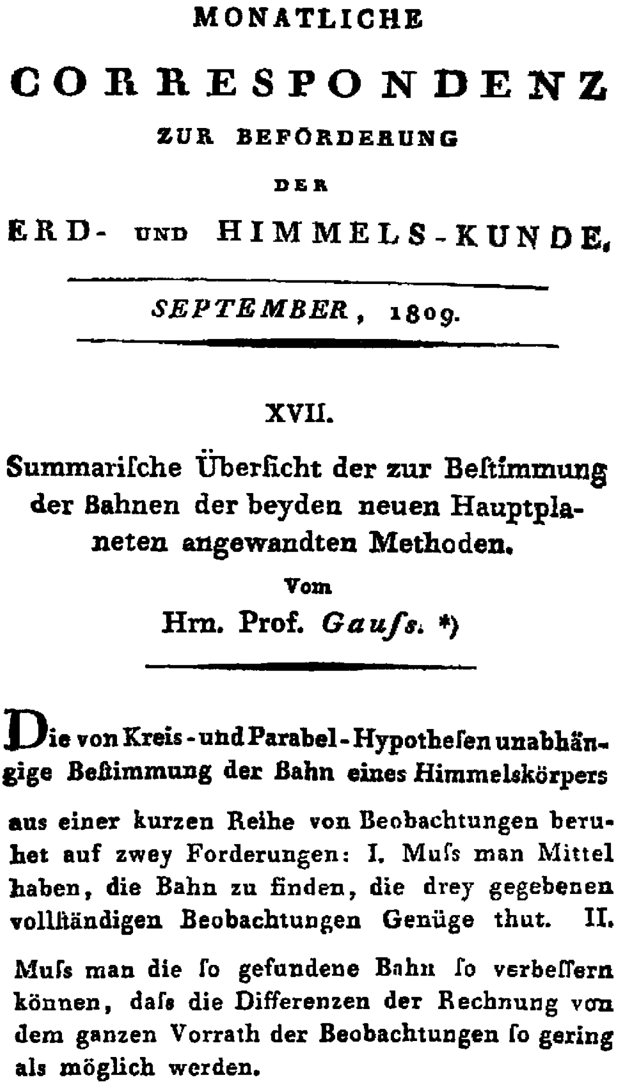
1. The problem of orbit determination shall be solved independently of hypotheses concerning the shape of the orbit. Gauss rejects in particular the assumption of a circular or a parabolic orbit (in the next section we will see that such assumptions considerably reduce the degree of difficulty of the problem).
2. In the same introductory sentence Gauss states that the problem has to be solved using a "short series of observations". The term is ambiguous, but the following developments document that "observations in a short time interval" (compared, e.g., to the revolution period of the observed body), are meant (as opposed to "few observations in an arbitrary time interval").
3. The problem shall be solved in two steps: In step "I." Gauss proposes to find a solution which represents three observations perfectly (residuals in the observed angles α and δ are zero).
4. In step "II." he proposes to improve the solution using in essence "his" method of least squares. This step is identical with the orbit improvement step specified in section 8.2.

Obviously, our treatment of the problem in section 8.2 is very closely related to Gauss' proposed procedure in 1809: The distinction is made between an initialization step and an orbit improvement step. The analogy goes even further:

The methods to be developed and discussed in this section will be mostly based on the assumption that the series of observations stems from a short time interval (it will be shown below that this condition can be released). It is thus fair to state that a modern treatment of the problem is (may be) still closely related to the principles published by Gauss in 1809.

It is interesting to note that step "I." of Gauss' procedure was harshly critizised in the 20st century by renowned practitioners of Celestial Mechanics. Paul Herget [54] writes, e.g., that *it would be a constructive achievement to dispel the myth that a "preliminary orbit can be computed from three observations"*. Observation errors, at times even gross errors (blunders), make such a procedure vulnerable. Taff [118] finds even stronger words.

One should not forget, however, that Gauss (as can be seen in the original text in Figure 8.2) simply viewed this first task as the initialization step, the result of which merely has to be of sufficiently high quality for the orbit improvement procedure to converge. Sometimes this original meaning is forgotten, probably because the first orbits obtained by Gauss' recipe often are of an amazingly high quality.



MONATLICHE

CORRESPONDENZ

ZUR BEFÖRDERUNG

DER

ERD- UND HIMMELS-KUNDE.

SEPTEMBER, 1809.

XVII.

Summariſche Überſicht der zur Beſtimmung der Bahnen der beyden neuen Hauptplaneten angewandten Methoden.

Vom

Hrn. Prof. *Gauſs.* *)

Die von Kreis- und Parabel-Hypotheſen unabhängige Beſtimmung der Bahn eines Himmelskörpers aus einer kurzen Reihe von Beobachtungen beruhet auf zwey Forderungen: I. Muſs man Mittel haben, die Bahn zu finden, die drey gegebenen vollſtändigen Beobachtungen Genüge thut. II. Muſs man die ſo gefundene Bahn ſo verbeſſern können, daſs die Differenzen der Rechnung von dem ganzen Vorrath der Beobachtungen ſo gering als möglich werden.

Fig. 8.2. The task of orbit determination according to Gauss

8.3.1 Determination of a Circular Orbit

If only two observations (astrometric positions) of a celestial body are available, it is not possible to determine all six elements of its orbit. The situation may arise in practice: After the discovery of a minor planet or a comet, the observer tries to make follow-up observations as soon as possible after the discovery. Using survey-type instruments (e.g., Schmidt-cameras) with fields of view of several degrees, it is usually possible to find the object after a few days without sophisticated prediction tools. If two observations are available, the apparent topocentric orbit may be approximated by a great circle on the celestial sphere assuming constant angular velocity. This approximation may be sufficient to follow the object's apparent trajectory over a few more days.

Figure II-4.47 tells that, statistically speaking, most minor planets are low eccentricity objects with eccentricities $e \leq 0.25$. One may therefore hope that a circular orbit is a rather good approximation of the true orbit. Experience tells, moreover, that observers are impatient and that they would like to gain insight into the characteristics of the orbit as soon as possible. The theoretician therefore usually has no choice, but to determine a circular orbit based on two astrometric positions given. It is perhaps not fair to make a general statement of this kind, but it is undoubtedly true if the observer's name is *Paul Wild*, and in our often rather long spells of clouded skies.

The assumption of a circular orbit reduces the number of unknowns from six to four, because the eccentricity is $e \overset{\text{def}}{=} 0$ and because the pericenter ω may be defined arbitrarily (e.g., as $\omega \overset{\text{def}}{=} 0°$). As each astrometric position provides two observation equations (one in α, one in δ, see (8.6)), it is theoretically possible to determine a circular orbit with two astrometric position. The parameters of the circular orbit are, e.g., the semi-major axis a, the inclination i, the longitude (or right ascension) of the ascending node Ω, and the node passing time T_0.

The attempt to determine a circular orbit may fail if the true eccentricity e of the object is sizeable, let us say $e > 0.1$, and if, by chance, the observations are near the pericenter or the apocenter. One might then change the rules to satisfy the observer's needs by determining an elliptic orbit with prescribed eccentricity e, and by assuming that the perigee or apogee lies in the middle between the two observation epochs. It is, however, much better to wait for additional observations .... If a solution is found, the four elements of a circular orbit are usually quite good approximations of the "true" values. A circular orbit may even be good enough to find the object in the next opposition.

Figure 8.3 illustrates the observation geometry of a circular orbit determination in the planetary system. The observer is supposed to be at positions E_i, at the observation times t_i , $i = 1, 2$, and the unit vectors $\boldsymbol{e}_i$ represent the astrometric position of the object at the observation times.

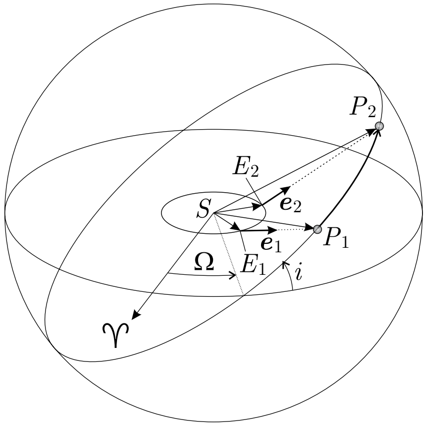


Fig. 8.3. Determination of a circular orbit using two astrometric positions $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ at observation times t_1 and t_2

The heliocentric positions P_i of the observed object at times $t_i - \Delta_i/c$, $i = 1, 2$, are the intersections of the straight lines defined by E_i and $\boldsymbol{e}_i$ with the sphere of radius a centered at the sun S. Denoting the heliocentric radius vectors of the observers E_i by $\boldsymbol{R}_i$, $i = 1, 2$, and those of the observed object P_i by $\boldsymbol{r}_i$, we obviously may write

$$\boldsymbol{r}_i = \boldsymbol{R}_i + \Delta_i \, \boldsymbol{e}_i \ , \quad i = 1, 2 \ , \tag{8.25}$$

where the topocentric distances Δ_i are obtained by squaring the above equation and by solving the resulting quadratic equation in Δ_i :

$$\Delta_i = - \boldsymbol{R}_i \cdot \boldsymbol{e}_i + \sqrt{(\boldsymbol{R}_i \cdot \boldsymbol{e}_i)^2 - (R_i^2 - a^2)} \ , \quad i = 1, 2 \ . \tag{8.26}$$

With the heliocentric vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ given, the heliocentric angle Δu_g between the two vectors may be calculated unambiguously – provided the observations stem from the same opposition – from Figure 8.3. The index g in the expression Δu_g stands for "geometrical method to calculate the heliocentric angle Δu between the two heliocentric position vectors of the celestial body". The same angle may also be calculated using the mean motion $n = \sqrt{\frac{k^2}{a^3}}$ of the object. In the case of a circular orbit this computation is particularly simple:

$$\Delta u_d = n \left[t_2 - t_1 - \left(\frac{\Delta_2}{c} - \frac{\Delta_1}{c} \right) \right] \ . \tag{8.27}$$

The index d in the expression Δu_d stands for "dynamical method to calculate the angle Δu". Both, Δu_d and Δu_g, are merely functions of the semi-major axis a. Obviously, we have to meet the condition

$$B(a) \overset{\text{def}}{=} \Delta u_g(a) - \Delta u_d(a) = 0 \tag{8.28}$$

for a circular orbit. This nonlinear equation in a may be solved iteratively, with the help of a graph of the function $B(a)$ within reasonable limits for a.

If a root is found, one obviously knows the corresponding semi-major axis a. Afterwards we have to calculate the remaining three elements i, Ω, and T_0. The Eulerian angles i and Ω of the orbital plane are obtained by computing the vector $\tilde{\boldsymbol{h}}$ normal to the orbital plane

$$\tilde{\boldsymbol{h}} = \boldsymbol{r}_1 \times \boldsymbol{r}_2 \tag{8.29}$$

and by using eqns. (4.5) thereafter to determine the elements i and Ω.

The argument of latitude u_1 (corresponding to the first observation time) is obtained by transforming the position vector $\boldsymbol{r}_1$ into the orbital system. The passing time through the ascending node is then given by

$$T_0 = t_1 - \frac{\Delta_1}{c} - \frac{u_1}{n} \; . \tag{8.30}$$

This concludes the determination of a circular orbit.

Let us now apply the theory to the observations of minor planet *Silentium*. Figure 8.4 shows the graph of function $B(a)$ in the range 0.975 AU $\leq a \leq$ 7.0 AU, when using the third and fifth observations in Table 8.1. The plot was generated with program ORBDET (see Chapter II- 8 of Part III). Obviously, there are multiple roots in the interval considered. The three solutions are made available to the program user (see Table 8.2) and the four orbital elements associated with the selected root are contained in the program output file. The orbital elements associated with each of the roots in Table 8.2 are listed in Table 8.3.

The first solution corresponds approximately to the orbit of the center of mass of the Earth-Moon system, which in turn coincides almost with the heliocentric trajectory of the observer. This type of solution is often encountered in the orbit determination process associated with objects in the planetary system and need not be further considered. In this case the first solution even might have been dismissed because the corresponding topocentric distances were negative (see Table 8.2).

The third solution at $a \approx 6.138$ AU would correspond to a retrograde orbit between Jupiter and Saturn, which renders this particular solution a rather unlikely candidate. The remaining second solution at $a \approx 2.377$ AU is in all likelihood the best approximation of the real orbit by a circular orbit (a guess which we will find confirmed later on).

Table 8.4 shows the residuals of all 1977 observations w.r.t. the second and third circular orbits of Table 8.3. It shows that the three observations (3-5) of September 18, 19, and 20 (actually contained in an interval only slightly

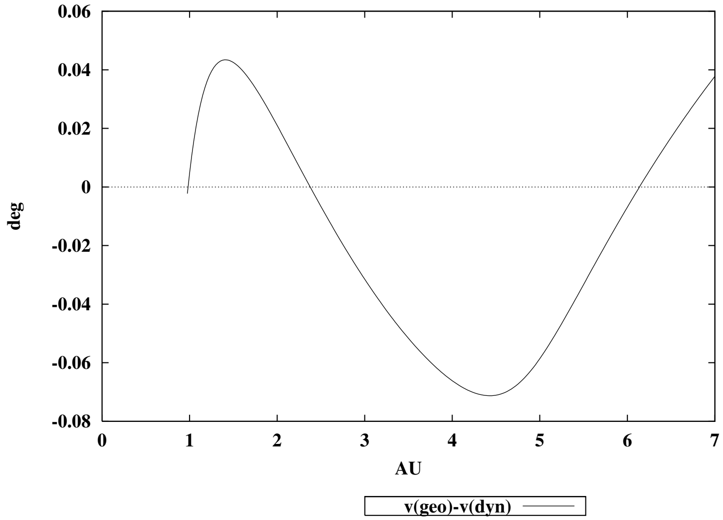


Fig. 8.4. Function $B(a) = \varDelta u_g - \varDelta u_d$ in circular orbit determination for minor planet *Silentium*

Table 8.2. Roots of circular orbit determination for minor planet *Silentium*

```
-----------------------------------------
NR       ROOT             TOPOC. DISTANCES
-----------------------------------------
 1      0.9827972      -0.01329  -0.01296
 2      2.3772021       1.38236   1.38250
 3      6.1381280       5.14382   5.14388
-----------------------------------------
  ROOT SELECTED:             2
```

Table 8.3. Different solutions of a first orbit determination using two observations

Element	Solution 1	Solution 2	Solution 3
Semi-major Axis a	0.982797 AU	2.377 AU	6.138 AU
Inclination i	0.045°	5.800°	145.380°
Long. of Asc. Node $\varOmega$	100.185°	6.955°	31.150°
Node Passing Time T_0	43508.541 days	43361.332 days	43358.112 days

Table 8.4. Residuals of the 1977 observations of *Silentium* w.r.t. the two circular orbits

Number	Solution 2		Solution 3	
	$\alpha \cos \delta$	δ	$\alpha \cos \delta$	δ
	$[\,''\,]$	$[\,''\,]$	$[\,''\,]$	$[\,''\,]$
1	−1308.13	−864.91	−4623.24	−2535.59
2	−937.84	−675.23	−3426.50	−1994.88
3	0.00	0.00	0.00	0.00
4	0.01	−1.74	0.64	−1.54
5	0.00	0.00	0.00	0.00
6	532.20	−2.96	1087.34	−37.02
7	702.35	13.99	1480.53	−2.83
8	1035.98	56.44	2280.27	94.14
9	1042.95	58.64	2266.47	91.17
10	1277.63	96.22	2849.74	179.78
11	1289.73	99.61	2879.32	185.69

longer than one day) are well represented by both solutions. Only when taking into account the observations $1-2$ and/or $6-11$ it becomes apparent that solution 2 is preferable. It is, by the way, interesting to note, that both solutions would be sufficient to find the minor planet within the time interval of roughly ± one month with a survey-telescope (remember that $3600''$ correspond to one degree).

8.3.2 The Two-Body Problem as a Boundary Value Problem

When solving the general orbit determination problem we have to determine eventually a set of orbital elements as stated in the original problem definition (8.1). In the problem of *first orbit determination* it may, however, be preferable to introduce auxiliary parameters for which good approximate values are easily available.

Figure 8.3 helps to explain the principle: Two observation epochs t_{b_1} and t_{b_2} , $b_i \in \{1, 2, \ldots, n\}$, $b_1 \neq b_2$, are selected as boundary epochs. The corresponding boundary vectors are written as:

$$\boldsymbol{r}_{b_i} = \boldsymbol{R}_{b_i} + \varDelta_{b_i}\, \boldsymbol{e}_{b_i} \ , \quad i = 1, 2 \ , \tag{8.31}$$

where $\boldsymbol{e}_{b_i}$ is defined by the right ascensions α_{b_i} and declinations δ_{b_i} . Using the set

$$\{p_1, p_2, \ldots, p_6\} \overset{\text{def}}{=} \{\varDelta_{b_1}, \varDelta_{b_2}, \alpha_{b_1}, \alpha_{b_2}, \delta_{b_1}, \delta_{b_2}\} \tag{8.32}$$

as auxiliary parameters in the first orbit determination process has the advantage that excellent approximations are available for the latter four parameters through the observations α'_{b_i} and δ'_{b_i} .

The selection (8.32) of auxiliary parameters implies that boundary value problems of the type

$$\begin{aligned} \ddot{\boldsymbol{r}} &= -\,\mu\,\frac{\boldsymbol{r}}{r^3} \\ \boldsymbol{r}\left(t_{b_1} - \frac{\Delta_{b_1}}{c}\right) &= \boldsymbol{r}_{b_1} \\ \boldsymbol{r}\left(t_{b_2} - \frac{\Delta_{b_2}}{c}\right) &= \boldsymbol{r}_{b_2} \end{aligned} \tag{8.33}$$

have to be solved.

If the two observation epochs t_{b_1} and t_{b_2} are close together, the problem (8.33) may be viewed and solved as a *local boundary value problem* in the sense of section 7.5.2. This has the distinct advantage that the perturbations might be included already in the process of first orbit determination and that the partial derivatives w.r.t. the parameters may be easily obtained by numerical methods, as well. The drawback must be seen in a restriction of the length of the time interval $|t_{b_2} - t_{b_1}|$: In principle one cannot guarantee that the interval is sufficiently short to allow the treatment of the problem (8.33) as a *local* boundary value problem *before* actually having solved it. Similar problems arise when solving the boundary value problem with the well-known classical tool of the *sector* : *triangle* – *ratios*, which we do not consider here (see, e.g., [114]).

Whereas such concerns usually are not justified when deriving a first orbit – it is in most cases possible to find two suitable boundary epochs – it is nevertheless comfortable that the above boundary value problem may in principle be solved without such restrictions. Let us now briefly sketch such a solution method.

In the analysis of the two-body problem in section 4.1 the initial value problem in the orbital plane was solved in two steps: The absolute value r of the radius vector $\boldsymbol{r}$ was derived as a function of the true anomaly (see eqn. (4.16))

$$r = \frac{p}{1 + e\,\cos v}\ , \tag{8.34}$$

then the differential equation (4.35) for the true anomaly was set up:

$$\dot{v} = \frac{h}{r^2} = \frac{\sqrt{\mu\,p}}{r^2} = \sqrt{\frac{\mu}{p^3}}\,\left(1 + e\,\cos v\right)^2\ . \tag{8.35}$$

Analytical solutions were given for elliptic, parabolic, and hyperbolic orbits in section 4.1.5.

The analytical solution for the length of the radius vector of the two-body problem may be used to find the solution of the boundary value problem (8.33), as well. From eqn. (8.34) we obtain for the boundary epochs:

$$\begin{aligned} e\,\cos v_{b_1} &= \frac{p}{r_{b_1}} - 1 \\ e\,\cos v_{b_2} &= \frac{p}{r_{b_2}} - 1 \ . \end{aligned} \tag{8.36}$$

By putting $v_{b_2} \overset{\text{def}}{=} v_{b_1} + \varDelta v$ (with $\varDelta v$ known as the heliocentric angle between the two boundary vectors), two equations for the determination of e and v_{b_1} are obtained, *provided* we consider the semi-latus rectum p of the conic section as known:

$$\begin{aligned} e\,\cos v_{b_1} &= \frac{p}{r_{b_1}} - 1 \\ &\overset{\text{def}}{=} A_1\,p - 1 \\ e\,\sin v_{b_1} &= \left\{ -\frac{p}{r_{b_2}} + 1 + \left(\frac{p}{r_{b_1}} - 1 \right) \cos \varDelta v \right\} \csc \varDelta v \\ &= \left\{ \left(\frac{\cos \varDelta v}{r_{b_1}} - \frac{1}{r_{b_2}} \right) \csc \varDelta v \right\} p \ + \ \{1 - \cos \varDelta v\} \csc \varDelta v \\ &\overset{\text{def}}{=} A_2\,p \ + B_2 \ . \end{aligned} \tag{8.37}$$

From the above equations the eccentricity e and the true anomaly v_{b_1} are obtained as

$$e^2 = \left(A_1^2 + A_2^2\right) p^2 \ - \ 2\left(A_1 - A_2\,B_2\right) p \ + \ 1 \ + \ B_2^2 \tag{8.38}$$

and

$$v_{b_1} = \arctan \left(\frac{A_2\,p + B_2}{A_1\,p - 1} \right) \ . \tag{8.39}$$

The semi-major axis a may then be obtained from the relations $p = a\,(1 - e^2)$ (for $e < 1$) and from $p = a\,(e^2 - 1)$ (for $e > 1$). The pericenter passing time T_0 is eventually obtained from the analytical solutions (4.56) (or from eqns. (4.55) if parabolic orbits shall be considered, as well) of the equation (8.35) of the anomaly.

From the construction of eqn. (8.38) we can see that a real eccentricity $e \geq 0$ results for each value of the semi-latus rectum p. Equation (8.38) may, however, also be used to define meaningful limits for the numerical values of the semi-latus rectum p of the conic section. The investigator usually is only interested in orbits with eccentricities between, let us say, $0 \leq e \leq 2$. The corresponding values for p may be obtained from eqn. (8.38) by writing it as a quadratic equation in the unknown p and the eccentricity e as a parameter. The solution of the resulting equation reads as:

$$p_{1,2}(e) = \frac{A_1 - A_2\,B_2 \pm \sqrt{(A_1 - A_2\,B_2)^2 - (A_1^2 + A_2^2)(1 - e^2 + B_2^2)}}{A_1^2 + A_2^2} \ . \tag{8.40}$$

Equation (8.40) tells that positive real values for the parameter will only result, if the determinant in the above equation is zero or positive:

$$\det\Big((A_1 - A_2 B_2)^2 - (A_1^2 + A_2^2)(1 - e^2 + B_2^2)\Big) \geq 0 \ . \qquad (8.41)$$

This condition leads to a lower limit for the eccentricity $e_{\min}$ of possible orbits (note that det $\rightarrow +\infty$ for $e \rightarrow \infty$ (because $A_1^2 + A_2^2 > 0$)). The upper limit is defined by the user's requirements (e.g., by $e_{\max} \leq 2$). For a given boundary value problem the semi-latus rectum p may therefore vary within the limits

$$p(e_{\min} \geq 0) \leq p \leq p(e_{\max}) \ . \qquad (8.42)$$

For each selected value for p in the range (8.42) we have constructed a conic section for which $r(v_{b_i}) = r_{b_i}$, $i = 1, 2$. From eqns. (8.38) and (8.39) the eccentricity e and the pericenter are known, as well (because the true anomaly v_{b_1} is known). The relationship between the true and eccentric anomalies and Kepler's equation (or the corresponding equations in the case of parabolic or hyperbolic orbits) may now be used to calculate the time T_0 of pericenter passage.

The analytical solution (4.56) of the equation (8.35) for the true anomaly v may be used to calculate the time difference between the observations. One just uses v_{b_1} and $v_{b_2} = v_{b_1} + \Delta v$ on the right-hand sides of eqns. (4.56), computes the corresponding values $t_{d_1}(p)$ and $t_{d_2}(p)$, and takes the difference $t_{d_2}(p) - t_{d_1}(p)$ of the two results. The index d stands, as usual, for the "dynamical" computation of the quantity associated with it. The correct value(s) of the semi-latus rectum p is (are) then obtained as the root(s) of the function

$$B(p) \overset{\text{def}}{=} t_{b_2} - t_{b_1} - \left(\frac{\Delta_{b_2}}{c} - \frac{\Delta_{b_1}}{c}\right) - \big(t_{d_2}(p) - t_{d_1}(p)\big) \ . \qquad (8.43)$$

The method outlined is in principle independent of the length of the time interval $[t_{b_2}, t_{b_1}]$. It is even capable of handling time intervals longer than one revolution period – provided the correct number of revolutions between the two boundary epochs may be assumed as known.

If Δv is a "reasonably small" angle (angles of $\Delta v \leq 60°$ usually pose no problem) the unknown parameter p is, however, best obtained by solving the boundary value problem (8.33) numerically and by computing p (and the other two-body elements of the orbit) from one of the sets of position- and velocity-vectors available after the solution of the problem (8.33). The computation of the semi-latus rectum p is handled in this way in program ORBDET (when selecting the solution option "boundary value"), where a collocation method of order $q = 12$ is used to solve the boundary value problem. It is, of course, very well possible to check a posteriori whether the solution of the boundary value problem was successful or not.

When setting up the observation equations we do not only have to solve the boundary value problem (8.33), we also need the partial derivatives of the reference orbit $\boldsymbol{r}(t)$ w.r.t. the orbit parameters p_i given in (8.32). This can be achieved in a very simple way: After having solved the boundary value problem (8.33) for the reference trajectory, we also know the orbital elements a, e, etc. In Chapter 5 we gave for each orbit type (ellipse, parabola, and hyperbola) formulae for the partial derivatives of the two-body orbit w.r.t. the orbital elements, allowing us to compute the partial derivatives of the reference orbit w.r.t. the classical orbital elements.

The six partial derivatives $\left(\frac{\partial \boldsymbol{r}}{\partial I_i}\right)(t)$ of the orbit $\boldsymbol{r}(t)$ w.r.t. the orbital elements $I_1 \overset{\text{def}}{=} p\,,\ I_2 \overset{\text{def}}{=} e\,,\ \ldots$ form a so-called complete system of solutions of the linear variational equations (5.22) associated with the two-body solution. This implies that any solution of the equations (5.22) may be written as a linear combination of the six partial derivatives defining the complete system of solutions. Let us designate the partial derivative of the orbit $\boldsymbol{r}(t)$ w.r.t. a general parameter p_i in the list (8.32) by

$$\boldsymbol{z}_i(t) \overset{\text{def}}{=} \left(\frac{\partial \boldsymbol{r}}{\partial p_i}\right)(t)\ , \quad i = 1, 2, \ldots, 6\ . \tag{8.44}$$

Writing these functions as linear combinations of the known partial derivatives w.r.t. the orbital elements leads to the representation

$$\boldsymbol{z}_i(t) = \sum_{l=1}^{6} c_{il} \left(\frac{\partial \boldsymbol{r}}{\partial I_l}\right)(t)\ , \tag{8.45}$$

where the coefficients c_{il} have to be determined from the boundary conditions, i.e., from

$$\boldsymbol{z}_i\left(t_{b_k} - \frac{\Delta_{b_k}}{c}\right) = \frac{\partial \boldsymbol{r}_{b_k}}{\partial p_i}\ , \quad k = 1, 2;\ i = 1, 2, \ldots, 6\ . \tag{8.46}$$

This implies that we have six linear equations for each of the partial derivatives for the determination of the six coefficients $c_{il}\,,\ l = 1, 2, \ldots, 6\,$, where the coefficient matrix is the same for all six partial derivatives, only the right-hand sides are different. These right-hand sides are computed as

$$
\begin{array}{lcl}
\dfrac{\partial \boldsymbol{r}_{b_1}}{\partial \mathit{\Delta}_{b_1}} = \boldsymbol{e}_{b_1} & ; & \dfrac{\partial \boldsymbol{r}_{b_2}}{\partial \mathit{\Delta}_{b_1}} = \boldsymbol{0} \\
\dfrac{\partial \boldsymbol{r}_{b_1}}{\partial \mathit{\Delta}_{b_2}} = \boldsymbol{0} & ; & \dfrac{\partial \boldsymbol{r}_{b_2}}{\partial \mathit{\Delta}_{b_2}} = \boldsymbol{e}_{b_2} \\
\dfrac{\partial \boldsymbol{r}_{b_1}}{\partial \alpha_{b_1}} = \begin{pmatrix} -\sin\alpha_{b_1}\cos\delta_{b_1} \\ +\cos\alpha_{b_1}\cos\delta_{b_1} \\ 0 \end{pmatrix} & ; & \dfrac{\partial \boldsymbol{r}_{b_2}}{\partial \alpha_{b_1}} = \boldsymbol{0} \\
\dfrac{\partial \boldsymbol{r}_{b_1}}{\partial \alpha_{b_2}} = \boldsymbol{0} & ; & \dfrac{\partial \boldsymbol{r}_{b_2}}{\partial \alpha_{b_2}} = \begin{pmatrix} -\sin\alpha_{b_2}\cos\delta_{b_2} \\ +\cos\alpha_{b_2}\cos\delta_{b_2} \\ 0 \end{pmatrix} \\
\dfrac{\partial \boldsymbol{r}_{b_1}}{\partial \delta_{b_1}} = \begin{pmatrix} -\cos\alpha_{b_1}\sin\delta_{b_1} \\ -\sin\alpha_{b_1}\sin\delta_{b_1} \\ +\cos\delta_{b_1} \end{pmatrix} & ; & \dfrac{\partial \boldsymbol{r}_{b_2}}{\partial \delta_{b_1}} = \boldsymbol{0} \\
\dfrac{\partial \boldsymbol{r}_{b_1}}{\partial \delta_{b_2}} = \boldsymbol{0} & ; & \dfrac{\partial \boldsymbol{r}_{b_2}}{\partial \delta_{b_2}} = \begin{pmatrix} -\cos\alpha_{b_2}\sin\delta_{b_2} \\ -\sin\alpha_{b_2}\sin\delta_{b_2} \\ +\cos\delta_{b_2} \end{pmatrix} .
\end{array}
\tag{8.47}
$$

The coefficients c_{il} of the partial derivatives (8.45) of the reference orbit w.r.t. the auxiliary parameters (8.32) are now easily computed by observing the condition equations (8.46) for each of the parameters using eqns. (8.47). As the left-hand sides of the condition equations are identical for all six parameters, only one matrix inversion is required for the required set of parameters.

It is obviously possible to change the parametrization from one set of six orbit parameters to another by making use of elementary properties of linear, homogeneous differential equations. It is in particular possible to use the set of parameters (8.32) defining a boundary value problem.

8.3.3 Orbit Determination as a Boundary Value Problem

Based on the principles outlined in the introductory section 8.2, based on the parametrization (8.32), and based on the analytical solution of the boundary value problem (and the associated variational equations) outlined in the preceding paragraph, the following robust orbit determination procedure may be defined:

- Two of n observation epochs t_{b_1} and t_{b_2} (corrected for the light-propagation times between the celestial body and the observer(s)) are selected as boundary epochs.
- The two boundary epochs may (but need not) be the first and last observation epochs. If the time interval containing the observations is short, it is best to select the first and last epoch; if the time interval is rather long, other selections may be better. The best performance may be expected if the difference $u_{b_2} - u_{b_1}$ between the arguments of latitude is between 10° and perhaps 30°.

- The boundary vectors $\boldsymbol{r}_{b_1}$ and $\boldsymbol{r}_{b_2}$ corresponding to the selected boundary epochs (corrected by the corresponding light travel time) are defined by

$$\boldsymbol{r}_{b_i} \overset{\text{def}}{=} \boldsymbol{R}_{b_i} + \Delta_{b_i}\, \boldsymbol{e}'_{b_i} \ , \quad i = 1, 2 \ , \tag{8.48}$$

  where $\boldsymbol{e}'_{b_i} = \boldsymbol{e}'_{b_i}(\alpha'_i, \delta'_i)$, $i = 1, 2$, are the observed unit vectors at observation times t_{b_1} and t_{b_2} .

- The topocentric distances Δ_{b_i} are not known. They may be systematically varied by assuming a scheme of the kind

$$\Delta_{b_1,k} \overset{\text{def}}{=} k\, d\Delta \, , \ k = 0, 1, 2, \ldots \qquad \text{and} \qquad \Delta_{b_2,j} \overset{\text{def}}{=} j\, d\Delta \, , \ j = 0, 1, 2, \ldots \ , \tag{8.49}$$

  where $d\Delta$ is a user-specified increment.

- The sum of the residuals squares in the minimum principle (8.8) is computed for each pair of topocentric distances according to

$$\Sigma(\Delta_{b_1}, \Delta_{b_2}) \overset{\text{def}}{=} \sum_{i=1}^{n} \left\{ \cos^2 \delta'_i \ v^2_{\alpha,i}(\Delta_{b_1}, \Delta_{b_2}) + v^2_{\delta,i}(\Delta_{b_1}, \Delta_{b_2}) \right\} \ , \tag{8.50}$$

  where the orbit used on the right-hand side to compute the residuals is defined by the parameters Δ_{b_1} , Δ_{b_2} , α'_{b_1} , α'_{b_2} , δ'_{b_1} , δ'_{b_2} .

- It is now a straightforward task to establish the minimum (or the minima) of the function $\Sigma(\Delta_{b_1}, \Delta_{b_2})$ in the two-dimensional table of values $\Sigma(\Delta_{b_1,k}, \Delta_{b_2,j})$.

- After having identified these minima of Σ-values, a conventional parameter estimation process including all six parameters (8.32) concludes the determination of the first orbit(s). Multiple solutions are possible.

The above procedure neither makes assumptions concerning the motion of the observed object nor concerning the spacing of the boundary epochs. The procedure promises to be robust and safe. This general procedure is, however, *not* implemented in program ORBDET, because in situations typical for first orbit determination, the two-dimensional search may be replaced by a more efficient one-dimensional search algorithm: If the time interval between the two boundary epochs is short compared to the expected revolution periods of the objects, the difference between the heliocentric distances r_{b_1} and r_{b_2} of the observed object at times t_{b_1} and t_{b_2} is comparatively small, as well. Having therefore assumed a value for the distance r_{b_1}, we also know a relatively good approximation for r_{b_2} by assuming simply $r_{b_2} \approx r_{b_1}$. This approximation may serve (exactly as in the case of the determination of a circular orbit) as an initial value for the true topocentric distance Δ_{b_2} , which then solves the quadratic equation:

$$\begin{aligned} r_{b_1}^2 &\approx r_{b_2}^2 \\ &= \left(\boldsymbol{R}_{b_2} + \Delta_{b_2}\, \boldsymbol{e}'_{b_2}\right)^2 \\ &= R_{b_2}^2 \,+\, 2\left(\boldsymbol{R}_{b_2} \cdot \boldsymbol{e}'_{b_2}\right)\Delta_{b_2} \,+\, \Delta_{b_2}^2 \ . \end{aligned} \tag{8.51}$$

The above quadratic equation in Δ_{b_2} has one positive solution if $R_{b_2} < r_{b_1}$; it has either no or two solutions if $R_{b_2} > r_{b_1}$.

This first approximation for Δ_{b_2} is iteratively improved by a conventional parameter estimation process with Δ_{b_2} as the only unknown parameter. Δ_{b_2} has to solve the linearized observation equations (8.6) (where the six orbit parameters we have only replaced by one, namely $I_1 \overset{\text{def}}{=} \Delta_{b_2}$).

The above considerations allow it to set up an algorithm requiring only a one-dimensional search over the topocentric distance referring to the first boundary epoch. The modified algorithm is implemented in program ORBDET and may be described as follows:

1. The topocentric distance Δ_{b_1} is systematically varied from zero to a user-defined upper limit in user-defined steps of $d\Delta$ (ORBDET proposes plausible values for planetary and satellite applications for the two user-defined quantities.)
2. For any given value of Δ_{b_1} one may compute the corresponding heliocentric (geocentric) distance at t_{b_1} by solving the equation in r_{b_1} :
$$r_{b_1}^2 = \left(\boldsymbol{R}_{b_1} + \Delta_{b_1}\, \boldsymbol{e}'_{b_1}\right)^2 \ . \tag{8.52}$$
3. With this value for r_{b_1} an approximate value for Δ_{b_2} follows by solving eqn. (8.51).
4. Keeping Δ_{b_1} and the angles α'_{b_i} and δ'_{b_i} , $i = 1, 2$, fixed, the linearized observation equations (8.6) are solved for the only remaining orbit parameter Δ_{b_2} .
5. The sum of residuals squares is computed for each tabular value Δ_{b_1}
$$\Sigma(\Delta_{b_1}) \overset{\text{def}}{=} \sum_{i=1}^{n} \left\{\cos^2 \delta'_i \; v_{\alpha_i}^2(\Delta_{b_1}) \,+\, v_{\delta_i}^2(\Delta_{b_1})\right\} \ . \tag{8.53}$$
6. In the (one-dimensional) table of values $\Sigma(\Delta_{b_1})$ the minimum (the minima) are established by numerical methods.
7. Having selected the minimum to be analyzed, a full orbit improvement step involving all six parameters (8.32) is performed. This terminates the initial orbit determination step.

This “one-dimensional” version of the first orbit determination problem is implemented in program ORBDET. It lists all minima found and asks the program operator to select the appropriate solution.

The method makes use of all observations available, and the result is the two-body orbit best representing the set of observations available (in the least-squares sense). An orbit improvement step using the conventional orbit parameters (following the first orbit determination outlined above) only is required, if perturbations are included.

8.3.4 Examples

The above theory is now applied to the observations of minor planet *Silentium.* Program ORBDET (see Chapter II- 8 of Part III) was used to produce Figure 8.5 showing the logarithm of the function

$$\sigma(\mathit{\Delta}_{b_1}) \overset{\text{def}}{=} \sqrt{\frac{\mathit{\Sigma}(\mathit{\Delta}_{b_1})}{n-1}} \tag{8.54}$$

of the topocentric distance $\mathit{\Delta}_{b_1}$. Program ORBDET departs in one minor point from the theory outlined above: Instead of using the function $\mathit{\Sigma}(\mathit{\Delta}_{b_1})$ standing for the sum of residuals squares, the function $\sigma(\mathit{\Delta}_{b_1})$ is shown, which stands for the mean error of the observations. For terrestrial photographic or CCD-observations using professional telescopes we expect $|\sigma(\mathit{\Delta}_{b_1})| \le 1''$ near the true solution. Figure 8.5 shows that our expectations are closely met in the case of minor planet *Silentium.*

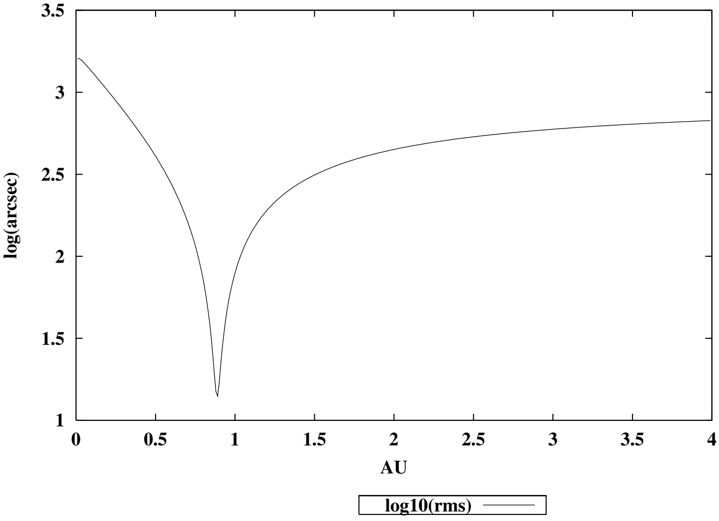


Fig. 8.5. Function $\sigma(\mathit{\Delta}_{b_1})$ of general orbit determination for minor planet *Silentium*

Observation epochs three and eleven (see Table 8.1) were used as boundary epochs. The function $\sigma(\varDelta_{b_1})$ in Figure 8.5 has exactly one minimum corresponding to a heliocentric distance of about $r_{b_1} \approx 1.84$ AU at the first boundary epoch. By linear interpolation in the table of $\sigma(r_{b_1})$-values, ORBDET determines a more precise location of the minimum $\varDelta_{b_1,\min}$ and the corresponding value $\varDelta_{b_2,\min}$ for $\varDelta_{b_2}$. The parameters

$$\{p_1, p_2, \ldots, p_6\} = \left\{\varDelta_{b_1,\min}, \varDelta_{b_2,\min}, \alpha'_{b_1}, \alpha'_{b_2}, \delta'_{b_1}, \alpha'_{b_2}\right\} \tag{8.55}$$

are then used as a priori values for an orbit determination process to determine the parameters (8.32).

Table 8.5 shows the residuals of the observations w.r.t. the best-fitting elliptic orbit. The table illustrates that observations may be marked (not used for the orbit determination process). Program ORBDET allows it also to remove previously introduced marks. Initially, only observations referring to one opposition are used.

If observations are marked or unmarked, program ORBDET either invokes an orbit improvement process without perturbations using the orbital elements p, e, i, $\varOmega$, ω, and T_0 (this particular set is valid for all possible conic sections) or an orbit improvement process with perturbations (to be addressed in section 8.4).

Table 8.5. Residuals of the 1977 observations of *Silentium* w.r.t. the two best-fitting elliptic orbits (all observations, all except marked observations)

Number	Solution (all)		Solution*	
	$\alpha \cos\delta$	δ	$\alpha \cos\delta$	δ
	[$''$]	[$''$]	[$''$]	[$''$]
1	−0.49	0.40	−0.41	0.99
2	0.77	−1.03	0.37	−0.78
3	0.31	−0.13	0.75	−0.43
4	−1.21	−0.28	−0.67	−0.81
5	−0.38	1.47	−0.17	0.94
6	3.67	1.10	4.51∗	0.97∗
7	−1.76	−0.29	−1.10	−0.40
8	−1.71	−2.59	−1.46∗	−2.70∗
9	1.04	−1.10	1.28	−1.22
10	−0.66	0.44	−0.74	0.30
11	0.43	1.65	0.34	1.51
Observations marked (*) Excluded				

Table 8.6 shows the orbital elements determined by program ORBDET corresponding to the first of the solutions in Table 8.5, together with the corresponding mean errors.

Table 8.6. First orbit determination for minor planet *Silentium* using all 1977 observations

Element	Value	Mean Error
Semi-latus rectum p	2.12633 AU	0.00064 AU
Semi-major Axis a	2.17577 AU	0.00064 AU
Eccentricity e	0.150735	0.000066
Inclination i	3.3739°	0.0024°
Long. of Asc. Node Ω	358.5430°	0.0119°
Arg. of Perihelion ω	1.3861°	0.0394°
Per. Passing Time T_0	43371.17 MJD	0.09 days

The above example is a routine case. It (probably) can be dealt with easily using any of the available orbit determination tools. We might, as a matter of fact, skip the very conservative and careful initialization process described above by defining the initial values for the topocentric distances at the boundary epochs "by experience" through $r_{b_1} \approx r_{b_2} \approx 2.7$ AU and by invoking the orbit improvement process with parameters (8.32) from the corresponding values of the topocentric distances (and the observed right ascensions and declinations). Such procedures were studied in [10]. They were found to be robust for orbits with reasonably small eccentricities.

Orbit determination does, however, not only know standard cases. Sometimes, the careful procedure developed above is actually required to study a particular orbit in detail. The orbit of comet *Panther* is a good example for a more demanding case.

Table 8.7 shows the (slightly rearranged) output of program ORBDET when determining the orbit of comet *Panther*, which was discovered in fall 1980, at Walgrave (England).

Orbit determination proved to be extremely delicate when only observations 1 to 8, spanning a time interval of about 11 days, were available. Using the orbit determination tools available in 1981 in Zimmerwald, it was close to impossible to determine a reliable orbit with these observations. This discouraging example led eventually to the development of the tools for first orbit determination outlined in this section.

Figure 8.6 shows the reason for the difficulties. Using observations 1 and 8 as boundary observations and the first eight observations to determine the orbit(s) led to the $\sigma(r_{b_1})$-function illustrated by this figure.

Table 8.7. Orbit determination for comet *Panther*

```
ORBIT DETERMINATION FOR OBJECT PANTHER   DATE: 02-DEC-01 TIME:  11:12
---------------------------------------------------------------------
OBSERVATIONS
------------
 NR   JJJJ MM DD.DDDDD  HH MM SS.SSS     VDD MM SS.SS     STA
  1   1980 12 27.76076   1 18 47 55.60    39 22 27.5     026
  2   1980 12 28.40625   1 18 48 14.47    39 32 01.3     026
  3   1980 12 28.72986   1 18 48 24.52    39 36 55.2     026
  4   1980 12 29.42986   1 18 48 45.24    39 47 40.5     026
  5   1980 12 30.95234   1 18 49 32.25    40 11 53.0     026
  6   1980 12 31.06597   1 18 49 35.87    40 13 52.4     026
  7   1980 12 31.74792   1 18 49 57.37    40 24 53.0     026
  8   1981  1  8.22194   1 18 54 10.61    42 44 30.3     026
  9   1981  1 23.77500   1 19 04 20.58    49 25 27.3     026
 10   1981  1 28.77222   1 19 07 59.99    52 11 39.3     026
 11   1981  4  9.06250   1 07 59 31.09    59 29 06.4     026

# OBS         =   8
RMS           = 5.37 " / 4.65"
TIME INTERVAL =   102.302 DAYS

Solution 1                                    Solution 2
**********                                    **********
P    = 1.855 AU +/- 0.182 AU       P    = 4.674 AU +/- 1.030 AU
A    = 2.221 AU +/- 0.356 AU       A    = 3.354 AU +/- 3.906 AU
E    =  0.406   +/- 0.064          E    = 1.547    +/- 0.425
I    =    74.08   +/- 2.31 DEG     I    =   85.46    +/- 1.40 DEG
NODE =   339.69   +/- 0.68 DEG     NODE =  326.95    +/- 3.29 DEG
PER  =   103.43   +/- 3.42 DEG     PER  =  108.67    +/- 2.74 DEG
TPER = 44618.44   +/- 2.46 DAYS    TPER =44641.13    +/- 0.86 DAYS

RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
                          Solution 1               Solution 2
                          **********               **********
  #      TIME        RA*COS(DE)      DE  MARK   RA*COS(DE)      DE   MARK
                        (")          (")           (")          (")
  1   44600.7613546     3.70        2.75            1.50        2.11
  2   44601.4068446    -5.03       -3.05           -4.46       -2.28
  3   44601.7304546     5.81       -0.77            4.53       -0.50
  4   44602.4304546    -4.41       -3.22           -3.28       -1.97
  5   44603.9529346    -0.75        1.73            0.11        0.94
  6   44604.0665647    -2.98        9.72           -1.13        8.97
  7   44604.7485147     3.36       -7.03            3.18       -7.47
  8   44612.2225349     0.32       -0.13           -0.45        0.21
  9   44627.7755952  -126.69     -273.45  *      -85.46      139.13    *
 10   44632.7728154  -248.21     -626.93  *     -135.86      254.63    *
 11   44703.0630971  4658.57    52066.37  *     2715.43   -10029.44    *
```

Instead of only one minimum, there are two relative minima. Their depths do not allow for a clear identification of the correct solution. Alternative orbits computed with program ORBDET, and starting from either of the two minima, differ significantly (see Table 8.7): in the first case (corresponding to the smaller heliocentric distance) an elliptic orbit with $a \approx 2.2$ AU and $e \approx 0.41$ is obtained. Neither the orbital elements nor the residuals can be used to rule out this solution. The second solution is hyperbolic. The resulting error is slightly better than in the first case, but based on the information available early in January 1980 it was not possible to rule out either of the two solutions. The "residuals" for observations 9 to 11 indicate that the second solution is nearer to the correct one. This is confirmed by Table 8.8, where all observations are used.

8.3.5 Determination of a Parabolic Orbit

In section 8.3.1 a method to determine a circular orbit was presented. In this case the number of orbit parameters was reduced from six to four. If

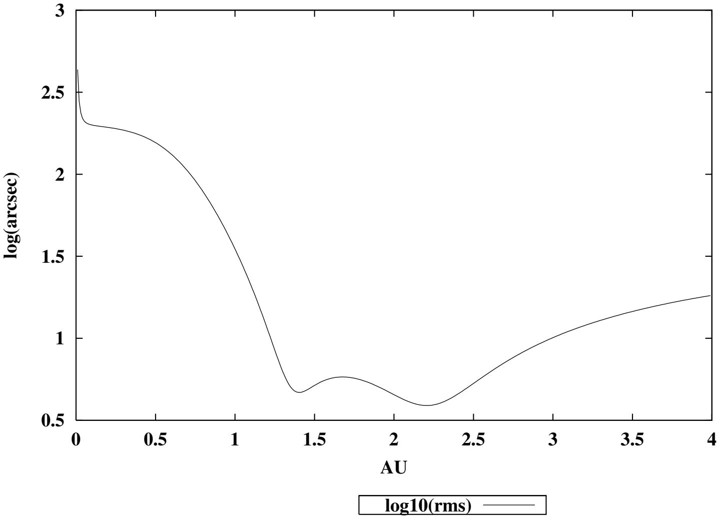


Fig. 8.6. Logarithm of function $\sigma(\varDelta_{b_1})$ for comet *Panther*

Table 8.8. Orbit determination for comet *Panther*

```
ORBIT DETERMINATION WITHOUT PERTURBATIONS
-----------------------------------------
# OBS         =  11
RMS           =     4.34 "
TIME INTERVAL =   102.302 DAYS
# ITERATIONS  =   2

P    =           3.3102 AU   +/-   0.0039 AU
A    =         729.5177 AU   +/- 574.7404 AU
E    =           0.9977      +/-   0.0018
I    =          82.642  DEG  +/-   0.010 DEG
NODE =         332.009  DEG  +/-   0.016 DEG
PER  =         105.593  DEG  +/-   0.023 DEG
TPER =       44631.304  DAYS +/-   0.046 DAYS

RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
  #       TIME          RA*COS(DE)      DE
                           (")          (")
  1   44600.7613546        3.69       -1.72
  2   44601.4068446       -3.73       -4.68
  3   44601.7304546        5.44       -1.94
  4   44602.4304546       -3.71       -2.11
  5   44603.9529346       -1.45        3.92
  6   44604.0665647       -3.03       12.07
  7   44604.7485147        1.44       -3.73
  8   44612.2225349        1.71       -0.95
  9   44627.7755952       -0.32       -1.95
 10   44632.7728154       -0.08        1.11
 11   44703.0630971       -0.04        0.00
```

an object in the planetary system is suspected to be a comet (not always a trivial decision), the hypothesis that the orbit is a parabola may make sense. In this case, the number of parameters is reduced from six to five, because the eccentricity is set to $e = 1$.

The first orbit determination method outlined in section 8.3.3 allows it to check the parabola hypothesis, as well: When calculating the sum of residuals squares $\Sigma(\Delta_{b_1})$, we obtain as a byproduct all six orbit parameters p, e, i, Ω, ω, and T_0 as a function of the heliocentric distance $r_{b_1} = |\boldsymbol{R}_{b_1} + \Delta_{b_1}\,\boldsymbol{e}_{b_1}|$. Figures 8.7 and 8.8 show the first two elements p and e when determining the first orbit of comet *Panther* using observations 1 to 8 in Table 8.7.

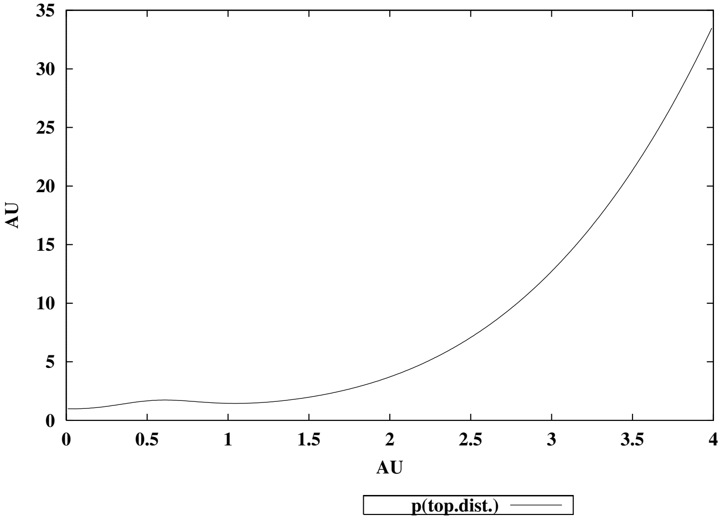


Fig. 8.7. Semi-latus rectum p for comet *Panther* as a function of topocentric distance Δ_{b_1}

Figure 8.8 is of particular interest: For $\Delta_{b_1} \approx 1.92$ the function $e(\Delta_{b_1})$ assumes the value $e \approx 1$, indicating that a parabolic solution might be a candidate orbit – in particular because the rms value (see Figure 8.6) associated with this topocentric distance is reasonably small.

Program ORBDET keeps track of plausible parabolic solutions. It produces a Table of parabolic orbits found together with the corresponding rms error and allows the program user to generate the best-fitting parabolic orbit instead of a general orbit. Table 8.9 shows the result of a parabolic orbit determination based on the first eight observations (see Table 8.7).

Clearly the parabolic orbit in Table 8.9 based on the observations 1 to 8 is much better (from the point of view of the mean errors of the elements, not

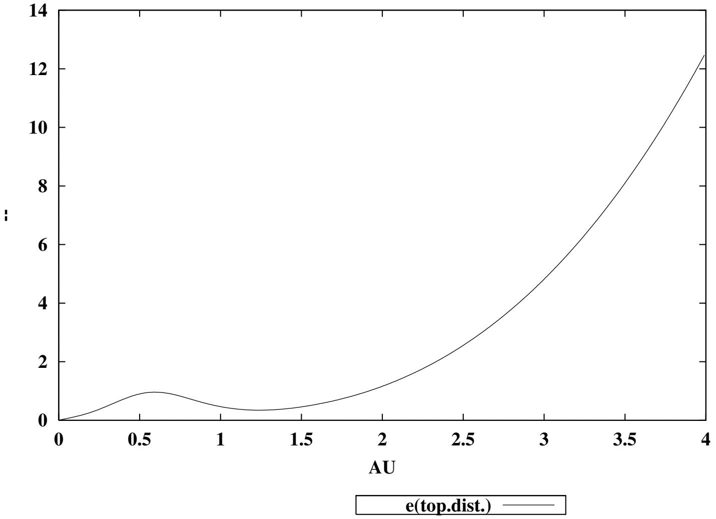


Fig. 8.8. Orbital eccentricity e of comet *Panther* as a function of topocentric distance Δ_{b_1}

Table 8.9. Orbit determination for comet *Panther*

```
ORBIT DETERMINATION FOR OBJECT PANTHER          DATE: 05-DEC-01 TIME:  18:39
-----------------------------------------------------------------------------
ORBIT DETERMINATION WITHOUT PERTURBATIONS
-----------------------------------------

# OBS         =   8
RMS           =      4.84 "
TIME INTERVAL =   102.302 DAYS
# ITERATIONS  =   1

P    =       3.350 AU   +/-  0.028 AU
I    =      82.870 DEG  +/-  0.173 DEG
NODE =     332.036 DEG  +/-  0.049 DEG
PER  =     104.480 DEG  +/-  0.935 DEG
TPER =   44630.220 DAYS +/-  0.962 DAYS

RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
  #      TIME          RA*COS(DE)      DE     MARK
                          (")          (")
  1   44600.7613546       4.64        1.43
  2   44601.4068446      -3.14       -2.84
  3   44601.7304546       5.81       -0.73
  4   44602.4304546      -3.67       -2.11
  5   44603.9529346      -2.08        1.70
  6   44604.0665647      -3.69        9.72
  7   44604.7485147       0.53       -6.78
  8   44612.2225349       1.62       -0.39
  9   44627.7755952      26.14      108.22     *
 10   44632.7728154      45.15      192.38     *
 11   44703.0630971    -881.39    -4347.64     *
```

of the residuals) than either the elliptic or the hyperbolic orbits in Table 8.7 based on the same observations, when estimating all six orbit parameters. The smaller errors in the parabolic elements are due to the fact that only five, and not six elements had to be determined (as in the case of the elliptic and hyperbolic orbit). A comparison of Tables 8.9 and 8.8 reveals that the parabolic orbit determined with eight observations stemming from a time interval of only ~ 11.5 days is quite close to the "true" orbit determined with all 11 observations stemming from a time interval of about 100 days.

8.3.6 Gaussian- vs. Laplacian-Type Orbit Determination

The algorithms for first orbit determination developed above followed in essential aspects the original Gaussian design as documented in Figure 8.2. We did in particular formulate and solve the problem as a boundary value problem by using the auxiliary variables (8.32). In other aspects we deviated considerably from the original Gaussian procedure. We did in particular *not* take over the Gaussian method to solve the boundary value problem (8.33).

Moreover, as opposed to the Gaussian procedure, we always used all observations contained in a reasonably short time interval, whereas Gauss took only three observations into account. This restriction allowed it to Gauss *not* to use the original observations α'_i and δ'_i but functions thereof which seemed more appropriate. The "original" Gaussian procedure will be briefly reviewed below.

Laplace developed a method which transformed the problem of first orbit determination at the earliest possible instant into an initial value problem. It is based on additional assumptions concerning the motion of the observer(s) and will be briefly addressed after the discussion of the Gaussian procedure.

Remarks concerning Gauss's "Original" Procedure. When restricting the number of astrometric position to three and when using the parameters (8.32) the number of unknowns is truly reduced from six to two because all the residuals will be zero (the number of unknowns equals the number of observations). Assuming that the first and the third observation epoch were used as boundary epochs, the six auxiliary parameters may be written as:

$$\{p_1, p_2, \ldots, p_6\} \overset{\text{def}}{=} \{\varDelta_1, \varDelta_3, \alpha'_1, \alpha'_3, \delta'_1, \delta'_3\} \ , \tag{8.56}$$

i.e., the latter four parameters are identical to the observed right ascensions and declinations. This is a consequence of the fact that the residuals of the observations must be zero, when the number of unknowns equals the number of observations.

It is thus only necessary to vary the remaining two unknowns $\varDelta_1$ and $\varDelta_3$ in such a way that the second observation also is exactly represented by the resulting orbit. As the residuals of this observation must be zero, as well, this latter condition is equivalent to the request that

$$\boldsymbol{R}_2 + \varDelta_2 \, \boldsymbol{e}_2' = \boldsymbol{r} \left(t_2 - \frac{\varDelta_2}{c} \right) \ , \tag{8.57}$$

where the (heliocentric resp. geocentric) position vector $\boldsymbol{R}_2$ of the observer and the unit vector $\boldsymbol{e}_2' = \boldsymbol{e}(\alpha_2', \delta_2')$ of the observed direction to the object are known.

Assuming for the moment, that we know the distances $\varDelta_i$, $i = 1, 2, 3$, we also know the boundary values and the boundary epochs:

$$\begin{aligned} \tilde{t}_1 = t_1 - \frac{\varDelta_1}{c} \quad &: \quad \boldsymbol{r}_1 = \boldsymbol{R}_1 + \varDelta_1 \, \boldsymbol{e}_1' \\ \tilde{t}_3 = t_3 - \frac{\varDelta_3}{c} \quad &: \quad \boldsymbol{r}_3 = \boldsymbol{R}_3 + \varDelta_3 \, \boldsymbol{e}_3' \ . \end{aligned} \tag{8.58}$$

It is well known that the two-body motion takes place in a plane. As we assume (at present) that the boundary vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$ are known, the orbital plane is known, as well. This means that the solution vector $\boldsymbol{r}(t)$ of the boundary value problem (8.33) may be written as

$$\boldsymbol{r}(t) = n_1(t) \, \boldsymbol{r}_1 + n_3(t) \, \boldsymbol{r}_3 \ . \tag{8.59}$$

By introducing the above formula into eqns. (8.33) we see that the functions n_1 and n_3 solve the following boundary value problem:

$$\begin{aligned} \ddot{n}_1 = - \mu \frac{n_1}{r^3} \quad &; \quad \ddot{n}_3 = - \mu \frac{n_3}{r^3} \\ n_1(\tilde{t}_1) = 1 \quad &; \quad n_1(\tilde{t}_3) = 0 \\ n_3(\tilde{t}_1) = 0 \quad &; \quad n_3(\tilde{t}_3) = 1 \ , \end{aligned} \tag{8.60}$$

where the heliocentric distance has to be calculated as $r = \sqrt{(n_1 \, \boldsymbol{r}_1 + n_3 \, \boldsymbol{r}_3)^2}$ in the above differential equations.

In all likelihood, the boundary value problem (8.60) qualifies as a *local boundary value problem*, which may be solved easily with collocation methods as developed in Chapter 7 – as long as the boundary values $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$ are known.

This statement allows us to consider the functions $n_1(t)$ and $n_3(t)$ as known functions of time. We may therefore in particular consider the function values at the time $\tilde{t}_2 = t_2 - \frac{\varDelta_2}{c}$ as known. Using the abbreviations

$$n_{12} \overset{\text{def}}{=} n_1(\tilde{t}_2) \quad ; \quad n_{32} \overset{\text{def}}{=} n_3(\tilde{t}_2) \tag{8.61}$$

we obtain what is often called the *fundamental equation of first orbit determination* by introducing formula (8.59) into the condition equation (8.57) to be met. Considering the construction (8.58) of the boundary vectors and the abbreviations (8.61) we obtain after a rearrangement of terms (collecting all contributions proportional to the topocentric distances on the left-hand side):

$$n_{12}\,\boldsymbol{e}_1'\,\varDelta_1\;-\;\boldsymbol{e}_2'\,\varDelta_2\;+\;n_{32}\,\boldsymbol{e}_3'\,\varDelta_3=-\;(n_{12}\,\boldsymbol{R}_1-\boldsymbol{R}_2+n_{32}\,\boldsymbol{R}_3)\;. \tag{8.62}$$

This *fundamental equation of orbit determination* is a system of three scalar equations. Obviously the true orbit must be a solution of it.

If n_{12} and n_{32} would not depend on the $\varDelta_i$, the fundamental equation could be viewed as a linear equation in the three unknown topocentric distances $\varDelta_i\,,\ i=1,2,3$. Unfortunately, the function values *do* depend on $\varDelta_1$ and $\varDelta_3$, because these quantities occur in the differential equation (see differential equation in boundary value problem (8.60)).

In view of the fact that the boundary values of the functions $n_1(t)$ and $n_2(t)$ do not depend on these distances, the dependence of n_{12} and n_{32} on these distances is rather weak. We may demonstrate in particular that these quantities are uniquely functions of r_2 (thus of $\varDelta_2$), by approximating the functions $n_1(t)$ and $n_3(t)$ as Taylor series of second order about t_2, the coefficients of which are determined in order to obey the boundary conditions of problem (8.60) and to satisfy the equations of motion of the same problem at $t=t_2$. (Subtleties like light propagation times are neglected in this approach):

$$\begin{aligned} n_1(t) &\overset{\text{def}}{=} n_{12}\;+\;(t-t_2)\,c_{11}\;+\;\tfrac{1}{2}\,(t-t_2)^2\,c_{12}\\ n_3(t) &\overset{\text{def}}{=} n_{32}\;+\;(t-t_2)\,c_{31}\;+\;\tfrac{1}{2}\,(t-t_2)^2\,c_{32}\;. \end{aligned} \tag{8.63}$$

The coefficients are determined through the following condition equations

$$\begin{aligned} n_1(t_1) &= n_{12}\;+\;(t_1-t_2)\,c_{11}\;+\;\tfrac{1}{2}\,(t_1-t_2)^2\,c_{12}=1\\ n_1(t_3) &= n_{12}\;+\;(t_3-t_2)\,c_{11}\;+\;\tfrac{1}{2}\,(t_3-t_2)^2\,c_{12}=0\\ \ddot{n}_1(t_2) &= c_{12}=-\,\frac{\mu}{r_2^3}\,n_{12}\\ n_3(t_1) &= n_{32}\;+\;(t_1-t_2)\,c_{31}\;+\;\tfrac{1}{2}\,(t_1-t_2)^2\,c_{32}=0\\ n_3(t_3) &= n_{32}\;+\;(t_3-t_2)\,c_{31}\;+\;\tfrac{1}{2}\,(t_3-t_2)^2\,c_{32}=1\\ \ddot{n}_3(t_2) &= c_{32}=-\,\frac{\mu}{r_2^3}\,n_{32}\;. \end{aligned} \tag{8.64}$$

In this approximation all coefficients of the Taylor series are functions of r_2. For the zero order terms (and only those are required subsequently) one obtains:

$$\begin{aligned} n_{12} &= + \frac{t_3 - t_2}{t_3 - t_1} \, \frac{1}{1 + \frac{\mu}{2\, r_2^3} \, (t_1 - t_2)(t_3 - t_2)} \\ n_{32} &= - \frac{t_1 - t_2}{t_3 - t_1} \, \frac{1}{1 + \frac{\mu}{2\, r_2^3} \, (t_1 - t_2)(t_3 - t_2)} \; . \end{aligned} \tag{8.65}$$

The above developments give rise to two methods for solving the fundamental equation (8.62):

Method 1: Direct Solution

1. A "plausible" value for r_2^3 is selected. (For applications in the planetary system one usually sets $r_2^3 = 20$, which is appropriate for minor planets between Mars and Jupiter).
2. Using this approximation, the function values n_{12} and n_{32} are computed according to formulae (8.65).
3. The fundamental equation (8.62) is solved for $\varDelta_1$, $\varDelta_2$, and $\varDelta_3$.
4. With the determined values $\varDelta_1$ and $\varDelta_3$ the boundary values $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$ are calculated using the relations (8.58).
5. The problem (8.60) is now solved as a local boundary value problem using, e.g., the collocation method developed in Chapter 7 or the analytical method developed in section 8.3.2.
6. Evaluating the solutions $n_1(t)$ and $n_2(t)$ at $t = \tilde{t}_2$ we obtain the best estimates for n_{12} and n_{32}.
7. The solution proceeds with step 3. until no further changes are observed in the topocentric distances $\varDelta_i$, $i = 1, 2, 3$.

Method 2: Search Method

1. The fundamental equation (8.62) is solved for a table of values $r_{2,k} \overset{\text{def}}{=} r_{2,\min} + k\, dr$, $k = 1, 2, \ldots$, where dr is a user-defined increment and $r_{2,\min}$ is the minimum value r_2 may assume.
2. The topocentric distance $\varDelta_{2,k}$ obtained as a solution of the fundamental equation (8.62) for the selection $r_{2,k}$ is used to compute the heliocentric (for satellite applications geocentric) distance using eqn. (8.57):
$$\tilde{r}_{2,k} \overset{\text{def}}{=} \sqrt{\left(\boldsymbol{R}_2^2 + \varDelta_{2,k}\, \boldsymbol{e}_2'\right)^2} \; . \tag{8.66}$$
3. The true (approximate) value for r_2 is obtained as the root of the function
$$B(r_2) \overset{\text{def}}{=} r_2 - \tilde{r}_2 = 0 \; . \tag{8.67}$$
4. Using the value of r_2 solving the equation $B(r_2) = 0$ as "plausible value", the orbit determination problem is iteratively improved with Method 1.

Method 1, if not preceded by a solution step using method 2, is a rather risky strategy. It usually works well for standard applications in the planetary system (i.e., for minor planets with orbits of rather small eccentricities between Mars and Jupiter). There are many variations of this method. Even Herget's generalization of the Gaussian method [54] must be viewed as a solution method of this type (apart from accepting more than three observations, the initialization differs by starting from two boundary vectors). In literature the boundary value problem (8.60) is usually solved by the well-established, but not easily understandable sector:triangle method (see, e.g., [114]). For cometary orbits this "shooting from the hips" method as reflected by Method 1 often fails.

Method 2 is related to our more general algorithms discussed above and implemented in program ORBDET. It has, however, the disadvantage that the orbit representation (Taylor series up to terms of second order) used for the search is rather poor. As mentioned, the Gaussian algorithms worked remarkably well for many (not to say most) applications in the planetary system.

When applying these methods to a broader class of objects (objects in the Edgeworth-Kuiper belt, to cometary orbits, NEA (Near Earth Asteroids)), and to artificial Earth satellites, a blind use of these methods may lead to a bad, at times catastrophic performance. This is probably the reason for the harsh criticism of the Gaussian methods in [118]. If the method is put on a rational basis and modernized as it was done in the preceding sections, Gaussian-type methods are among the best available.

Let us mention that in Method 2 the fundamental equations (8.62) might be reduced to one scalar equation in $\varDelta_2$ by formally eliminating the first and third topocentric distances because n_{12} and n_{32} are only functions of the parameter r_2^3, which in turn is a function of $\varDelta_2$.

First Orbit Determination According to Laplace. Easy understandability is a particularly attractive aspect of Laplacian-type orbit determination methods. The method, attributed to Laplace, makes the attempt to transform a problem, which from its nature is a boundary value problem, into an initial value problem.

This may be achieved by fitting the original observations, e.g., the right ascensions and declinations α_i', $i = 1, 2, \ldots, n$, and δ_i', $i = 1, 2, \ldots, n$, by Taylor series of a certain order $2 \le q \le n-1$ in time:

$$\alpha(t) = \sum_{i=0}^{q} \frac{1}{i!}\, \alpha_0^i \, (t-t_0)^i \quad ; \quad \delta(t) = \sum_{i=0}^{q} \frac{1}{i!}\, \delta_0^i \, (t-t_0)^i \ , \quad 2 \le q \le n-1 \ . \tag{8.68}$$

The derivatives α_0^i and δ_0^i of the observations referring to an arbitrary development epoch t_0 (best selected in the middle of the interval covered by observations) are established by a conventional least-squares procedure. From now on, only the coefficients of the Taylor series (8.68), but no longer the observations themselves, are analyzed. As the entire information is now related to one epoch t_0, one may truly speak of an initial value problem.

Using the representation (8.68) for the observation, we may also represent the observed unit vector $\boldsymbol{e}(t)$

$$\boldsymbol{e}(t) \overset{\text{def}}{=} \boldsymbol{e}\big(\alpha(t), \delta(t)\big) \tag{8.69}$$

and its derivatives as a function of time. The first and the second derivative are computed as

$$\dot{\boldsymbol{e}}(t) = \left(\frac{\partial \boldsymbol{e}}{\partial \alpha}\right)(t)\,\dot{\alpha}(t) + \left(\frac{\partial \boldsymbol{e}}{\partial \delta}\right)(t)\,\dot{\delta}(t) \tag{8.70}$$

and

$$\begin{aligned}\ddot{\boldsymbol{e}}(t) = \left(\frac{\partial \boldsymbol{e}}{\partial \alpha}\right)(t)\,\ddot{\alpha}(t) &+ \left(\frac{\partial \boldsymbol{e}}{\partial \delta}\right)(t)\,\ddot{\delta}(t) + \left(\frac{\partial^2 \boldsymbol{e}}{\partial \alpha^2}\right)(t)\,\dot{\alpha}^2(t) \\ +\,2\left(\frac{\partial^2 \boldsymbol{e}}{\partial \alpha\,\partial \delta}\right)(t)\,\dot{\alpha}(t)\dot{\delta}(t) &+ \left(\frac{\partial^2 \boldsymbol{e}}{\partial \delta^2}\right)(t)\,\dot{\delta}^2(t)\ ,\end{aligned} \tag{8.71}$$

where the derivatives of the unit vector w.r.t. α and δ are obtained from the definition of the unit vector in the equatorial system

$$\boldsymbol{e} = \begin{pmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{pmatrix} .$$

We may in particular consider the unit vector and its first two derivatives w.r.t. time at t_0 as known quantities.

For applications in the planetary system it is assumed in Laplacian-type orbit determination procedures that not only the observed celestial body, but also the observer is moving on a conic section around the Sun:

$$\begin{aligned}\ddot{\boldsymbol{r}} &= -\,k^2\,\frac{\boldsymbol{r}}{r^3} \\ \ddot{\boldsymbol{R}} &= -\,k^2\,(1 + m_{♁})\,\frac{\boldsymbol{R}}{R^3}\ .\end{aligned} \tag{8.72}$$

For applications to artificial Earth satellites the latter assumption does not make sense. One may assume, however, that the observer moves on a known analytical trajectory, the first two derivatives thereof are known, as well. Taff [118] discusses generalizations of this kind.

Assuming that the trajectory of the observer is known and that the motion of the observed object is defined by the differential equation of the two-body problem we may derive an equation of motion of the observed body w.r.t. the observer by taking the second time derivative of

$$\boldsymbol{r}(t) = \boldsymbol{R}(t) + \mathit{\Delta}(t)\, \boldsymbol{e}(t) \; . \tag{8.73}$$

For applications in the planetary system one obtains for time $t = t_0$, where $\boldsymbol{e}_0 \overset{\text{def}}{=} \boldsymbol{e}(t_0)$, etc.,

$$\boldsymbol{e}_0\, \ddot{\mathit{\Delta}}_0 + 2\, \dot{\boldsymbol{e}}_0\, \dot{\mathit{\Delta}}_0 + \left(\ddot{\boldsymbol{e}}_0 + k^2\, \frac{\boldsymbol{e}_0}{r_0^3} \right) \mathit{\Delta}_0 = -\, k^2 \left(\frac{1}{r_0^3} - \frac{1}{R_0^3} \right) \boldsymbol{R}_0 \tag{8.74}$$

and for artificial satellites (see, e.g., [118]):

$$\boldsymbol{e}_0\, \ddot{\mathit{\Delta}}_0 + 2\dot{\boldsymbol{e}}_0\, \dot{\mathit{\Delta}}_0 + \left(\ddot{\boldsymbol{e}}_0 + GM\, \frac{\boldsymbol{e}_0}{r_0^3} \right) \mathit{\Delta}_0 = -\, GM\, \frac{\boldsymbol{R}_0}{r_0^3} - \ddot{\boldsymbol{R}}_0 \; . \tag{8.75}$$

Equations (8.74) and (8.75), respectively, are the fundamental equations for Laplacian-type orbit determination. They correspond to the fundamental equation (8.62) of Gaussian-type first orbit determination.

Equations (8.74) and (8.75), respectively, are non-linear equations in the three scalar unknowns $\mathit{\Delta}_0$, $\dot{\mathit{\Delta}}_0$, and $\ddot{\mathit{\Delta}}_0$, because the only other quantity r_0 which is unknown in these equations may be written as a function of $\mathit{\Delta}_0$:

$$r_0^2 = \left(\boldsymbol{R}_0 + \mathit{\Delta}_0\, \boldsymbol{e}_0 \right)^2 \; . \tag{8.76}$$

In practice, the systems of equations (8.74) and (8.75) may, e.g., be solved by formally eliminating $\dot{\mathit{\Delta}}_0$ and $\ddot{\mathit{\Delta}}_0$ using basic methods of linear algebra. The result consists of an equation containing only the unknown quantities $\mathit{\Delta}_0$ and r_0, which, in view of eqn. (8.76), means that one scalar equation for the unknown $\mathit{\Delta}_0$ is obtained. The resulting equation may even be transformed into an equation of degree eight in the remaining unknown (see, e.g., [114]) then solved by numerical methods.

Equations (8.74) and (8.75) may also be solved without such algebraic transformations. Note, that the equations would be linear in the three scalar quantities $\mathit{\Delta}_0$, $\dot{\mathit{\Delta}}_0$, and $\ddot{\mathit{\Delta}}_0$, if r_0 were known. Therefore, one may solve these equations for a table of values $r_{0,k} = r_{0,\min} + k\, dr$, $k = 1, 2, \ldots$, determine the corresponding values $\mathit{\Delta}_{0,k}$ by formally solving the equations (8.74) or (8.75) for the three unknowns (where the derivatives of $\mathit{\Delta}_0$ are of no interest when determining the correct value $\mathit{\Delta}_0$), and determine the root of the function

$$B(r_0) \overset{\text{def}}{=} \sqrt{(\boldsymbol{R}_0 + \mathit{\Delta}_0\; \boldsymbol{e}_0)^2} - r_0 = 0 \tag{8.77}$$

with numerical methods using the table of values $r_{0,k}$ and $\mathit{\Delta}_{0,k}$.

Let us assume now that we have found the correct root(s) $\mathit{\Delta}_0$ of the function (8.77). In order to obtain the initial position- and velocity-vector at time t_0, we need not only $\mathit{\Delta}_0$ but also the corresponding value $\dot{\mathit{\Delta}}_0$, which is obtained by solving the fundamental equations (8.74) or (8.75) for the value r_0 corresponding to the root(s) of equation (8.77). The position- and velocity- vectors at t_0 are then obtained by:

$$\begin{aligned} \boldsymbol{r}(t_0) &= \boldsymbol{R}_0 + \Delta_0\, \boldsymbol{e}_0 \\ \dot{\boldsymbol{r}}(t_0) &= \dot{\boldsymbol{R}}_0 + \dot{\Delta}_0\, \boldsymbol{e}_0 + \Delta_0\, \dot{\boldsymbol{e}}_0 \ . \end{aligned} \tag{8.78}$$

The light-propagation time was neglected in our review of the essentials of Laplacian-type orbit determination procedures. If one would wish to implement such procedures, this aspect should be handled correctly. Approximately, we just might identify the starting epoch with $t_0 - \frac{\Delta_0}{c}$ and not with t_0 .

Classical Gaussian vs. Laplacian Procedures. In the preceding paragraphs we have analyzed the classical Gaussian and Laplacian orbit determination methods. In both cases *fundamental equations* of a very similar structure were obtained.

It is often viewed as an advantage of the Laplacian-type orbit determination that implicitly all observations are used in the first orbit determination process by fitting the observations by a development of type (8.68). This is true if the observations are very noisy, and if many (dozens) of observations are available in a short time interval. It is, however, questionable whether this should be viewed as an advantage of the Laplacian method. One might as well generate three artificial observations for three different epochs using the same representation (8.68) and then apply the Gaussian procedure.

The advantage of the Laplacian procedure may also be viewed as its disadvantage: Fitting the observations by polynomials only makes sense if the trajectories of the observer(s) and the celestial body are smooth enough to justify such fits. For applications to artificial satellites this argument strictly rules out the use of optical observations stemming from different observatories, whereas this is possible without problems for Gaussian-type algorithms. When applied to minor planets or comets the same argument is not valid. In a fair approximation all observatories on the surface of the Earth may be identified with the center of mass of the Earth-Moon system. The error introduced through such neglects depend on the distance of the celestial object. In most cases Laplacian first orbit determination is still possible – naturally with a slightly reduced accuracy. It should be pointed out, however, that no such concerns exist for Gaussian procedures.

The main advantage of Laplace-type orbit determination algorithms resides in their transparency. The main advantage of Gaussian-type orbit determination procedures resides in their mathematical correctness. They are based on the minimum number of assumptions. The algorithms for first orbit determination developed in this Chapter and implemented in program ORBDET are based on the original Gaussian ideas as reflected by Figure 8.2. ORBDET was successfully applied to all kinds of minor planets ("ordinary" objects between Mars and Jupiter, objects in the Edgeworth-Kuiper belt, NEAs, artificial Earth satellites, and space debris). The search table in ORBDET is optimized for “normal” objects. When dealing with objects in the Edgeworth-Kuiper belt, the table has to be extended toward larger geocentric distances (see section II- 8).

8.4 Orbit Improvement: Examples

The theory of orbit improvement is in essence identical with the theory of nonlinear parameter estimation. It is based on the three keywords *initialization*, *iteration*, and *termination*. This theory – using pure orbit determination as an example – was already outlined in the paragraph 8.2.1. The delicate part of the orbit determination problem, namely the initialization of the iterative orbit improvement process, was discussed in detail in the previous section 8.3, which is why we may confine ourselves here with presenting and discussing a few examples.

Let us first study the orbit of a minor planet using observations from several oppositions. After a successful first orbit determination procedure using the observations of only one opposition, observations from different oppositions (planetary system) shall be analyzed now. When determining the best possible two-body orbit from observation series covering more than one revolution period, one usually realizes that these observations are not well represented. This behavior is illustrated by Table 8.10, where the two-body orbit of minor planet *Silentium* is determined not only with the observations of the 1977 opposition, but also with the Zimmerwald observations of the 1993 opposition.

Table 8.11 provides another example using observations made at Zimmerwald. Observations of this kind are, e.g., obtained during search campaigns for space debris. Figure 8.9 illustrates two possibilities to observe geostationary satellites. In Figure 8.9 (left) the satellites are moving w.r.t. the star background, because the telescope is compensated for the diurnal motion of the stars, in Figure 8.9 (right) the telescope is in the staring mode, i.e., it is pointing into one and the same Earth-fixed direction; this mode implies that geostationary objects are mapped as “points” and the stars as dashes. For space debris search campaigns in the geostationary belt the second mode is preferable.

In such campaigns an observation session may last for several hours (or even the entire night). The geostationary belt is systematically screened for known and unknown objects. In this observation mode, the same objects are observed a few times on successive CCD-frames in a short observation interval of a few minutes only (usually 30 seconds to three minutes), then (depending on the search strategy) again after a longer time interval of one to a few hours. In the example documented by Table 8.11 observations of the object

Table 8.10. Orbit elements of Silentium, from the 1977 and 1993 observations, and residuals w.r.t. the best-fitting two-body orbits

```
ORBIT DETERMINATION WITHOUT PERTURBATIONS
-----------------------------------------
# OBS        =  15
RMS          =     14.61 "
TIME INTERVAL =  5866.909 DAYS
# ITERATIONS  =   5
P    =        2.1264637123 +/-       0.0004464957 AU
A    =        2.1768293257 AU
E    =        0.1521089870 +/-       0.0006633385
I    =           3.3608431 +/-          0.0027896 (DEG)
NODE =         358.3983727 +/-          0.0216815 (DEG)
PER  =           0.7998919 +/-          0.4220762 (DEG)
TPER =       43369.5325464 +/-          0.9338195 DAYS

RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
  #       TIME         RA*COS(DE)      DE    MARK
                          (")          (")
  1   43405.0530984      11.62      -28.67
  2   43408.0443285      13.34      -29.09
  3   43434.9175693      17.07       -9.84
  4   43435.9880594      14.84       -9.06
  5   43436.0380594      15.21       -7.06
  6   43450.7821598       2.79        1.46
  7   43452.8172299      -5.37        1.23
  8   43456.0026400     -11.18        1.06
  9   43456.0457000      -9.16        2.67
 10   43457.9149701     -12.61        4.55
 11   43458.0116701     -12.67        6.11
 12   49249.9211816      -7.90        5.20
 13   49250.9444416      -4.84        6.89
 14   49270.9250022      -8.48       24.60
 15   49271.9625022      -9.37       24.67
```

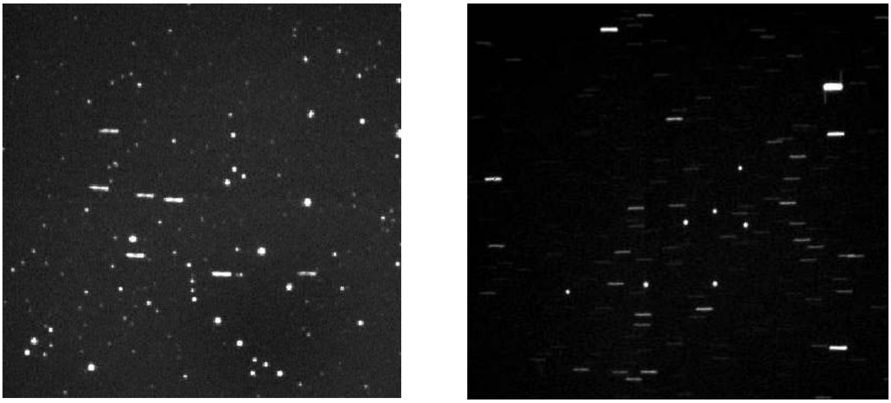


Fig. 8.9. Astrometric CCD observation of geostationary satellites with the Zimmerwald observatory (left: sidereal tracking, right: staring mode, i.e., tracking turned off)

with COSPAR number 97049B (MeteoSat 7), stemming from two successive nights (January 2/3 and January 3/4, 2002), are analyzed.

In Table 8.11 a circular orbit was determined first from observations 1 and 4, separated by $1^{\mathrm{m}}30^{\mathrm{s}}$. The search for space debris in the geostationary belt provides, by the way, an excellent example for the usefulness of determining a circular orbit. In view of the fact that objects in the geostationary belt usually are in orbits of samll eccentricities, a circular orbit determined with two observations separated by only 1-2 minutes is usually of better quality than an elliptic orbit based on all observations contained in the same short time interval. Also, when the search strategy is optimized to obtain the maximum number of new objects, usually only two observations are obtained of the same object in the short time interval mentioned.

After the determination of a circular orbit, all 31 observations are used to determine the best-possible two-body orbit. Obviously, the residuals show a systematic behavior which is about one order of magnitude above the measurement noise (which is of the order of about $0.10 - 0.20''$ per observation, either $\alpha \cos\delta$ or δ).

Program ORBDET allows to considerably improve the force field for applications in the planetary system and in satellite geodesy. The improvement implies that the equations of motion have to be modified and solved with alternative methods.

For the purpose of the improvement of planetary orbits the equations of motion (3.21) are implemented in program ORBDET:

$$\ddot{\boldsymbol{r}} = -\,k^2 \,\frac{\boldsymbol{r}}{r^3} \;-\; k^2 \sum_{j=1}^{n} m_j \left\{ \frac{\boldsymbol{r} - \boldsymbol{r}_j}{|\boldsymbol{r} - \boldsymbol{r}_j|^3} + \frac{\boldsymbol{r}_j}{r_j^3} \right\} \;, \tag{8.79}$$

where the mass m_0 of the central body (i.e., the Sun) was set to $m_0 = 1$. A (sub)set of the nine major planets may be selected as perturbing bodies by the user. The position vectors of the selected planets are computed approximately using the formulae contained in Meeus [72].

The variational equations (see Chapter 5 of Part III) have to be integrated simultaneously with the equations of motion. The variational equations associated with the two-body problem were derived in Chapter 5, eqn. (5.22). The variational equations corresponding to the equations of motion (8.79) read as

$$\ddot{\boldsymbol{z}} = \mathbf{A}\,\boldsymbol{z} \;, \tag{8.80}$$

where $\boldsymbol{z}$ stands for the partial derivative of the celestial body's position vector w.r.t. one of the six orbit parameters, and where

$$\mathbf{A} = -\frac{k^2}{r^3}\left\{\mathbf{E} - \frac{3}{r^2}\,\boldsymbol{r} \otimes \boldsymbol{r}^T\right\} - k^2 \sum_{j=1}^{n} \frac{m_j}{|\boldsymbol{r} - \boldsymbol{r}_j|^3}\left\{\mathbf{E} - \frac{3\,(\boldsymbol{r} - \boldsymbol{r}_j) \otimes (\boldsymbol{r} - \boldsymbol{r}_j)^T}{(\boldsymbol{r} - \boldsymbol{r}_j)^2}\right\} . \tag{8.81}$$

Table 8.11. Residuals and orbital elements of *MeteoSat 7* (97049B) w.r.t. the two best-fitting two-body orbits

```
ORBIT DETERMINATION FOR OBJECT 97049B        DATE: 17-FEB-02 TIME:  07:31
--------------------------------------------------------------------------
 NR   OBJECT      YYYY MM DD   HH MM SS.SSS      RA(H)        DE(DEG)
   1  97049B      2002  1  2   18 48 19.841     1.593343     -6.92939
   2  97049B      2002  1  2   18 48 49.853     1.601697     -6.92921
   3  97049B      2002  1  2   18 49 19.852     1.610052     -6.92888
   4  97049B      2002  1  2   18 49 49.862     1.618411     -6.92847
   5  97049B      2002  1  2   19 16  2.389     2.056358     -6.91162
   6  97049B      2002  1  2   19 16 32.410     2.064720     -6.91134
   7  97049B      2002  1  2   19 17  2.410     2.073070     -6.91098
   8  97049B      2002  1  2   19 17 32.421     2.081431     -6.91073
   9  97049B      2002  1  2   19 44 57.938     2.539669     -6.89405
  10  97049B      2002  1  2   20 43 39.847     3.520331     -6.86279
  11  97049B      2002  1  2   20 44  9.888     3.528693     -6.86258
  12  97049B      2002  1  2   20 44 41.276     3.537434     -6.86235
  13  97049B      2002  1  2   20 45 15.526     3.546970     -6.86207
  14  97049B      2002  1  2   21 27 51.772     4.258681     -6.84455
  15  97049B      2002  1  2   21 28 21.812     4.267043     -6.84437
  16  97049B      2002  1  2   21 28 51.813     4.275396     -6.84412
  17  97049B      2002  1  2   21 29 21.803     4.283751     -6.84406
  18  97049B      2002  1  3    1 27  7.210     8.254993     -6.84110
  19  97049B      2002  1  3    1 27 37.221     8.263351     -6.84127
  20  97049B      2002  1  3    1 28 16.438     8.274271     -6.84154
  21  97049B      2002  1  3   17 59 29.162     0.840787     -6.96331
  22  97049B      2002  1  3   17 59 59.192     0.849155     -6.96313
  23  97049B      2002  1  3   18  0 29.203     0.857514     -6.96278
  24  97049B      2002  1  3   18  0 59.194     0.865863     -6.96239
  25  97049B      2002  1  3   19 48 34.731     2.663822     -6.89356
  26  97049B      2002  1  3   19 49  4.781     2.672189     -6.89338
  27  97049B      2002  1  3   19 49 34.772     2.680544     -6.89311
  28  97049B      2002  1  3   19 50  4.773     2.688902     -6.89275
  29  97049B      2002  1  3   22 20 17.917     5.198412     -6.83207
  30  97049B      2002  1  3   22 20 47.963     5.206773     -6.83185
  31  97049B      2002  1  3   22 21 17.264     5.214928     -6.83161

CIRCULAR ORBIT FOR OBJECT 97049B           DATE: 17-FEB-02 TIME:  07:31
------------------------------------------------------------------------
   RESIDUALS IN ARCSECONDS             ELEMENTS
   -----------------------             --------

    I      RA        DE
    1      0.00      0.00     A    =   42172088.0
    2     -0.17     -0.48     E    =     0.000000
    3     -0.11     -0.38     I    =     0.136461
    4      0.02      0.00     NODE =     5.101649
    5      3.73      4.07  *  PER  =     0.000000
    6      3.87      4.03  *  TPER =    -4694.839
    7      3.68      4.27  *
    ..     ....      ....
    ..     ....      ....

ORBIT DETERMINATION WITHOUT PERTURBATIONS
-----------------------------------------
# OBS         =  31
RMS           =      4.77 "
TIME INTERVAL = 99177.422 SEC
# ITERATIONS  =  11

P    =     42166181.332356 +/-          36.080678 M
A    =     42166183.757693 +/-          36.071062 M
E    =        0.0002398302 +/-       0.0000086390
I    =           0.1436683 +/-          0.0004173 (DEG)
NODE =           6.3765383 +/-          0.1427097 (DEG)
PER  =         -75.6944880 +/-          2.1518521 (DEG)
TPER =      -22500.4164587 +/-        507.3420231 SEC
```

Table 8.11. (Continued.)

```
RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
  #       TIME          RA*COS(DE)     DE    MARK
          (s)              (")          (")
  1        0.0000000       1.85        5.80
  2       30.0110397       1.60        5.24
  3       60.0108481       1.57        5.26
  4       90.0210240       1.61        5.56
  5     1662.5476799       1.42        5.31
  6     1692.5682240       1.50        5.19
  7     1722.5688955       1.25        5.35
  8     1752.5790720       1.40        5.12
  9     3398.0964480       0.83        5.17
 10     6920.0058239      -1.15        5.24
 11     6950.0462400      -1.28        5.12
 12     6981.4344958      -1.24        5.07
 13     7015.6843195      -1.24        5.12
 14     9571.9302718      -2.19        4.12
 15     9601.9706880      -2.27        4.12
 16     9631.9713595      -2.28        4.36
 17     9661.9616640      -2.00        3.95
 18    23927.3688959       0.06        0.84
 19    23957.3790719       0.26        0.81
 20    23996.5968956       0.41        0.64
 21    83469.3206399      -2.26       -6.55
 22    83499.3506879      -2.12       -7.12
 23    83529.3617276      -2.11       -7.10
 24    83559.3520315      -2.36       -6.94
 25    90014.8895998       1.01       -8.79
 26    90044.9395198       1.00       -9.19
 27    90074.9306875       1.17       -9.25
 28    90104.9313596       1.34       -9.00
 29    99118.0756799       1.56       -9.52
 30    99148.1212799       1.39       -9.09
 31    99177.4221117       1.24       -8.59
```

All in all six variational equations, corresponding to the six orbit parameters (initial conditions or osculating orbital elements) have to be integrated. In order to avoid case-distinctions and in order to avoid delicate special cases which might give rise to singularities (e.g., starting a general orbit determination procedure from a circular orbit), program ORBDET uses internally the rectangular components of the position- and the velocity-vectors as orbit parameters for orbit improvement with perturbations. The results and the associated mean errors are, however, transformed into osculating orbital elements at the initial epoch after the completion of the orbit improvement procedure.

Table 8.12 shows the osculating elements of Silentium's orbit at the initial epoch (first of the 1977 Zimmerwald observations) and the residuals w.r.t. the best-fitting orbit, where the perturbations due to the planets Mars, Jupiter, Saturn, and Neptune were included.

The residuals are (more or less) randomly distributed. When comparing the results in Tables 8.10 and 8.12 one notes that the mean error of the semi-major axis is much smaller in the latter case than expected according to a $1/\sqrt{n}$-law (n being the number of observations). This is due to the circumstance that the semi-major axis a defines the mean motion. The mean error of the semi-major axis thus decreases not only according to the $1/\sqrt{n}$-law but it is also proportional to the length of the time interval covered by the observations.

Table 8.12. Orbit elements of *Silentium*, from the 1977 and 1993 observations, and residuals w.r.t. best-fitting perturbed orbit

```
ORBIT DETERMINATION WITH PERTURBATIONS
--------------------------------------
# OBS        =  14
RMS          =      1.34 "
TIME INTERVAL =  5866.909 DAYS
# ITERATIONS  =   3

P    =        2.1271958752 +/-       0.0000405400 AU
A    =        2.1765687310 +/-       0.0000013513 AU
E    =        0.1506114297 +/-       0.0000598213
I    =           3.3772449 +/-          0.0002634 (DEG)
NODE =          -1.4361728 +/-          0.0019823 (DEG)
PER  =           1.4341112 +/-          0.0397738 (DEG)
TPER =       43371.2849103 +/-          0.0887213 DAYS

RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
  #       TIME         RA*COS(DE)      DE    MARK
                          (")          (")
  1   43405.0530984      -1.55        2.18
  2   43408.0443285      -0.41        0.69
  3   43434.9175693       1.14        0.35
  4   43435.9880594      -0.47        0.23
  5   43436.0380594      -0.08        2.19
  6   43450.7821598       4.09        0.30     *
  7   43452.8172299      -0.82       -1.01
  8   43456.0026400      -1.24       -2.67
  9   43456.0457000       0.85       -1.08
 10   43457.9149701       0.70        0.02
 11   43458.0116701       0.81        1.55
 12   49249.9211816      -0.85       -1.95
 13   49250.9444416       1.72       -1.19
 14   49270.9250022       0.20        0.61
 15   49271.9625022       0.10        0.07
```

For satellite geodetic applications, program ORBDET may take into account the perturbations due to the Earth's oblateness and due to the gravitational attraction exerted by Sun and Moon. The equations of motion assume the form (compare eqns. (3.143) and (II- 3.9)):

$$\ddot{\boldsymbol{r}} = + \frac{1}{r^3} \left\{ - GM\,\boldsymbol{r} + \frac{3}{2}\,\frac{\tilde{C}_{20}}{r^2} \begin{pmatrix} r_1 \left(1 - 5\,\dfrac{r_3^2}{r^2}\right) \\ r_2 \left(1 - 5\,\dfrac{r_3^2}{r^2}\right) \\ r_3 \left(3 - 5\,\dfrac{r_3^2}{r^2}\right) \end{pmatrix} \right\} \tag{8.82}$$
$$- GM_{☾} \left\{ \frac{\boldsymbol{r} - \boldsymbol{r}_{☾}}{|\boldsymbol{r} - \boldsymbol{r}_{☾}|^3} + \frac{\boldsymbol{r}_{☾}}{r_{☾}^3} \right\} - GM_{\odot} \left\{ \frac{\boldsymbol{r} - \boldsymbol{r}_{\odot}}{|\boldsymbol{r} - \boldsymbol{r}_{\odot}|^3} + \frac{\boldsymbol{r}_{\odot}}{r_{\odot}^3} \right\} .$$

These equations of motion are hardly able to model the orbit of a LEO over a time interval of more than a few revolutions. For satellites in the geostationary belt and in the height of the current generation of navigation satellites, the model is good enough to represent direction observations with an accuracy well below the arcsecond. Better modelling capabilities will be provided and discussed below.

As in the case of planetary and cometary orbits, the variational equations are solved simultaneously with the primary equations (8.82) when the orbit of an artificial satellite or of a space debris is determined. The structure of the variational equations is the same as in the case of the planetary system, i.e., the variational equations are given by eqns. (8.80). The matrix **A** may be written as a sum of the terms due to the point mass contributions from Earth, Moon, and Sun, and due to the Earth's oblateness J_{20}. The former three terms may be taken over from eqn. (8.81). Thus, only the contribution due to C_{20} would have to be addressed here, but the exercise is left to the readers.

Table 8.13 shows the result of an orbit improvement process including the oblateness and the third-body perturbations due to the Moon and the Sun using all observations of object 97049B in the nights of January 2/3 and 3/4. When comparing Tables 8.13 and 8.11 we may obviously conclude that the improved force model is much better suited to represent the observations. Small systematic effects of the order of a few $0.10''$ exist and will be addressed below.

In program ORBDET the equations of motion and the associated variational equations are solved with the numerical methods developed in Chapter 7. As

Table 8.13. Residuals and orbital elements of observations of *MeteoSat 7* (97049B) w.r.t. the best-fitting perturbed orbit

```
ORBIT DETERMINATION WITH PERTURBATIONS
--------------------------------------
# OBS         =  31
RMS           =       0.30 "
TIME INTERVAL = 99177.422 SEC
# ITERATIONS  =   3

P    =      42167030.659765 +/-           2.237931 M
A    =      42167033.622328 +/-           2.237217 M
E    =         0.0002650619 +/-       0.0000005579
I    =            0.1441395 +/-          0.0000258 (DEG)
NODE =            5.8957628 +/-          0.0088765 (DEG)
PER  =          -72.2644793 +/-          0.1161062 (DEG)
TPER =       -21794.5172848 +/-         27.3039082 SEC

RESIDUALS IN RIGHT ASCENSION AND DECLINATION
--------------------------------------------
  #       TIME         RA*COS(DE)      DE    MARK
                          (")          (")
  1       0.0000000       0.45        0.46
  2      30.0110397       0.20       -0.11
  3      60.0108481       0.17       -0.10
  4      90.0210240       0.22        0.20
  5    1662.5476799       0.30       -0.10
  6    1692.5682240       0.39       -0.22
  7    1722.5688955       0.14       -0.06
  8    1752.5790720       0.31       -0.29
  9    3398.0964480       0.22       -0.19
 10    6920.0058239      -0.33        0.32
 11    6950.0462400      -0.45        0.20
 12    6981.4344958      -0.40        0.16
 13    7015.6843195      -0.39        0.22
 14    9571.9302718      -0.23       -0.22
 15    9601.9706880      -0.30       -0.22
 16    9631.9713595      -0.30        0.03
 17    9661.9616640      -0.01       -0.37
 18   23927.3688959      -0.16        0.40
 19   23957.3790719       0.03        0.38
 20   23996.5968956       0.14        0.21
 21   83469.3206399      -0.34        0.54
 22   83499.3506879      -0.22       -0.02
 23   83529.3617276      -0.23        0.02
 24   83559.3520315      -0.50        0.18
 25   90014.8895998      -0.02        0.30
 26   90044.9395198      -0.04       -0.09
 27   90074.9306875       0.12       -0.15
 28   90104.9313596       0.29        0.11
 29   99118.0756799       0.48       -0.72
 30   99148.1212799       0.30       -0.30
 31   99177.4221117       0.16        0.18
```

the program should be able to accommodate any orbit types (not only orbits with small eccentricities), the integration method of choice is the collocation method with stepsize control. The numerical integration is set up in such a way that the initial epoch may be located anywhere w.r.t. the observation epoch. As a matter of fact the initial epoch is automatically set to coincide with the first observation epoch selected for first orbit determination. This means that forward and backward integration is possible.

Program ORBDET, as provided on the CD accompanying this book, is a tool for first orbit determination (using all observations of one opposition resp. one satellite pass) and for orbit improvement for observation series spanning at maximum a few revolution periods of the observed object.

For orbit improvement based on longer time spans of observations, more elaborate tools for orbit modelling have to be used. The included software package contains program SATORB (see Chapter II- 7 of Part III), which was already extensively used in Chapter II- 3 to illustrate the perturbations acting on artificial satellites. Program SATORB may not only be used as a tool to produce satellite ephemerides, but also as an orbit determination tool. SATORB will be used in section 8.5.4 to produce ephemerides of LEOs, for which geocentric position vectors (and possibly position differences) were obtained from spaceborne GPS receivers. Here we use it to generate an improved orbit for a time series of approximately 11 days of observations of Meteosat 7. The first two days of observations are identical with those in Table 8.11.

Table 8.14 shows the results of the orbit improvement process. The program needs approximate orbital elements. These were taken over from program

Table 8.14. Orbital elements of *MeteoSat 7* (97049B), and residuals w.r.t. the best-fitting perturbed orbit (150 observations in 11 days, January 2 - 13, 2002)

```
ORBIT DETERMINATION FOR OBJECT 97049B           DATE: 17-FEB-02 TIME:  10:45
----------------------------------------------------------------------------

ORBIT DETERMINATION USING *.OBS-FILES FOR   1 SATELLITE(S)
**********************************************************
SATELLITE   1    ARC          =        1
                 FROM (MJD)   = 52276.784
                 TO (MJD)     = 52287.787
                 # OBS-EPOCHS =       155
                 # ITERATIONS =         5
-----------------------------------------

ORBITAL ELEMENTS AND THEIR RMS ERRORS
***************************************************
OSCULATION EPOCH = 52276.7835630 MJD
SEMIMAJOR AXIS   =  42167095.628 M   +-       0.195 M
REV. PERIOD U    =      1436.218 MIN
ECCENTRICITY     =  0.0002625494 --- +-0.0000000667
INCLINATION      =     0.1441769 DEG +- 0.000007362
R.A. OF NODE     =     5.9114545 DEG +- 0.002311926
ARG OF PERIGEE   =   -72.6985835 DEG +- 0.045279186
ARG OF LAT AT TO =    18.7996063 DEG +- 0.002311457
***************************************************

NUMBER OF DYNAMICAL PARAMETERS   :  1
DECOMPOSITION TYPE (1=RSW, 2=DYX):  2
********************************************************
PARAMETER = D0 VALUE =-.666859D-07 +-0.763417D-09 M/S**2
********************************************************
SAT   1 : RMS=   0.21"  # OBS =  310 # PARMS =  7 BETA= -22.74 DEG
```

ORBDET (not documented in Table 8.14). The best available force model was used for the program run: The Earth's potential coefficients up to terms of degree and order 70 were used, gravitational attractions from Sun and Moon were included, the solid Earth tides and the relativistic corrections were taken into account, as well. Seven orbit parameters, namely the osculating elements at the initial epoch and a constant acceleration due to the direct radiation pressure term (D_0) were estimated. Unnecessary to say that the orbital elements are much better determined from this long series of observations than for the series documented in Table 8.13. One can also see from Table 8.14 that the direct radiation pressure, $D_0 \approx 6.67 \cdot 10^{-8}$ m/s^2, is significant and has to be solved for. If only the six initial elements are determined, the rms error of the single observation grows from $0.21''$ to $1.05''$.

Figure 8.10 shows the residuals of the observations in $\alpha \cos \delta$ and in δ as a function of time (left) and as a function of the observation number. The figure to the left nicely illustrates the problematic of observing merely from one observatory: Due to the fact that optical observations can be made at night only, the distribution of the observations in the time interval between January 3 and January 13, 2002, is far from homogeneous. Also, it is hard to judge whether or not there are small systematic errors left in the residuals. Figure 8.10 indicates that this is not the case. A much better distribution of observations could be achieved with $3 - 6$ observatories well distributed in geographical longitude.

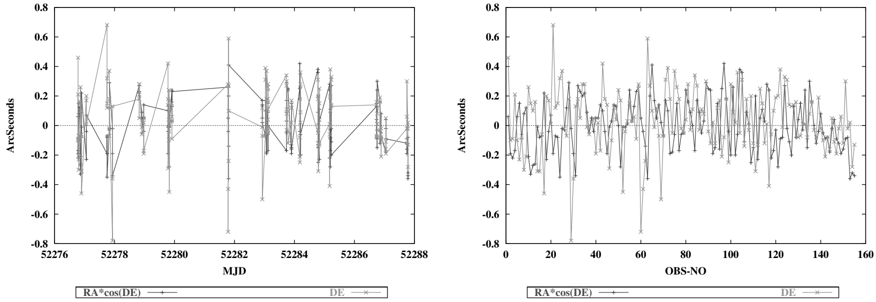


Fig. 8.10. Residuals of orbit improvement using 150 observations in 11 days of Meteosat 7 (97049B), as a function of time (left) and of observation number (right)

8.5 Parameter Estimation in Satellite Geodesy

The development of space geodesy, with today's cornerstones SLR/LLR and GPS, was outlined in section 2.2 together with that of the third space geodetic technique, namely VLBI. The three techniques are fundamental for the establishment, maintenance, and continued refinement of the celestial and the terrestrial reference systems. *Nolens-volens* non-geometrical parameters related to the Earth's atmosphere have to be determined (or eliminated) as well, because the atmosphere influences the propagation characteristics of the electromagnetic signals observed by the space geodetic techniques.

The theory of parameter estimation is that of non-linear parameter estimation – assuming that approximate values for all parameters estimated are available. This theory was already outlined in paragraph 8.2.1 using astrometric observations as an example. To cope with more general problems, we simply have to replace the observation equations (8.6) for the astrometric observations α_i' and δ_i' by the observation equations for the functions o_i' accessible to the satellite geodetic techniques mentioned above. Moreover we have to acknowledge that the main difference of the general parameter estimation problem in satellite geodesy w.r.t. the orbit improvement problems treated so far resides in the much enlarged parameter space containing many more different parameter types. The orbit parameters usually are only a small subset of these parameters.

Subsequently only SLR/LLR and GPS, as satellite geodetic techniques, are considered in this chapter, which is devoted to orbit determination and parameter estimation. The "orbits" of Quasars are not interesting in this context. This does not mean that the VLBI technique is uninteresting for Earth sciences: The discussions in Chapter II- 2 will show that VLBI is of fundamental nature, in particular for the establishment of the celestial reference system and for modelling precession and nutation.

Rather complex parameter estimations are performed routinely in satellite geodesy. It is the purpose of this section to give some insight into the work performed by the space geodetic services ILRS and IGS (see Table 2.5).

8.5.1 The General Task

The parameters accessible to satellite geodetic techniques are:

- orbit parameters defining the initial state vector of a satellite at one particular epoch t_0 and the dynamical parameters characterizing the force model necessary to describe the orbital motion of the satellites,
- coordinates of the observing sites in an Earth-fixed system (if the rigid body model is used for the Earth) or in a Tissérand system (see section 3.3.7), if non-rigid Earth models are used,
- motion of the observatories relative to a Tissérand system,
- Earth rotation and Earth orientation parameters defining the transformation between the Earth-fixed and the inertial system at the measurement epoch (in particular x and y, the polar wobble coordinates, the length of day, and possibly nutation terms),

- atmosphere parameters defining the tropospheric refraction correction,
- atmosphere parameters defining the ionospheric refraction correction, as well as
- technique-specific parameters (e.g., station-specific range biases, satellite clock parameter, receiver clock parameter, etc.).

It may be possible to determine some of the parameters with additional and independent measurements with sufficient precision. This may considerably reduce the dimension of the parameter space. The parameter space underlines the interdisciplinary potential of satellite geodesy.

The *orbital arc* or simply the *arc* as a contiguous, limited part of the satellite's trajectory plays an important role in satellite geodesy. Whereas usually all observations made of a minor planet or a comet are represented eventually by one contiguous trajectory described by one set of initial conditions (in general only six parameters), the same is usually not true in satellite geodesy. When a satellite is observed over many years, it is not possible to describe the entire time period covered by observations by one set of initial conditions and the appropriate dynamical parameters. In such cases the orbital arc may be broken up into arcs of lengths defined by the analyst. Each arc is described by exactly one initial state vector and the dynamical parameters. This is why usually many sets of parameters defining the initial values have to be estimated in satellite geodetic analyses. Alternatively (in an attempt to avoid breaking up a trajectory into orbital arcs) one might introduce so-called pseudo-stochastic parameters (fudge parameters) or replace the deterministic differential equation for the trajectory by a stochastic differential equation, which in essence asks for the replacement of least-squares estimators by digital filters (the topic will be briefly addressed in section 8.5.4).

The orbit parameters (and the satellite ephemerides derived thereof) are of different importance for different applications. Whereas a satellite ephemeris usually is a necessary, but otherwise unimportant by-product when analyzing the orbits of SLR satellites, the GPS ephemerides are of central importance. It is only these high-precision ephemerides which allow it subsequently to a broad (unlimited) user community to perform navigation or positioning tasks with an accuracy in essence only limited by (a) the accuracy of the determined orbits and (b) by tropospheric refraction effects.

8.5.2 Satellite Laser Ranging

Within each of the satellite geodetic observation techniques the satellites are observed from sites on the Earth's surface (or from low orbiting satellites (LEOs)) at well-defined epochs. The definition of the observed quantity, the *observable*, is different for each observation technique. So far we uniquely dealt with astrometric positions as observed quantities. Omitting technique-specific parameters (like range or clock biases) the observation equations for *Satellite Laser Ranging (SLR)* read as:

$$\begin{aligned}\Delta\tilde{t} = {} & \frac{1}{c}\left\{ \left| \boldsymbol{r}(t+\Delta t_1) - \boldsymbol{R}(t) \right| + \left| \boldsymbol{R}(t+\Delta t_1+\Delta t_2) - \boldsymbol{r}(t+\Delta t_1) \right| \right\} \\ & + 2\,\Delta t_{\mathrm{trop}} \ ,\end{aligned} \qquad (8.83)$$

where

- t is the observation epoch, defined here as the epoch, when the Laser pulse left the observatory,
- c is the speed of light in vacuum,
- $\boldsymbol{R}(t)$ is the geocentric position vector of the observatory,
- $\boldsymbol{r}(t)$ that of the satellite at time t,
- Δt_1 is the propagation time of the Laser pulse from the observatory to the satellite,
- Δt_2 the light propagation time of the pulse reflected at the satellite back to the observatory,
- Δt_{trop} is the tropospheric refraction correction, and
- $\Delta\tilde{t}$ is the observable, the signal propagation time from the observatory to the satellite and back.

The position vectors $\boldsymbol{R}(t)$ and $\boldsymbol{r}(t)$ have to be expressed in the same system, either in the Earth-fixed or in the inertial system.

One may assume that the station $\boldsymbol{R}(t)$ is in rectilinear motion with constant velocity during the short time interval $[t, t+\Delta t_1+\Delta t_2]$, which is why the above equation may be simplified to read as:

$$\Delta\tilde{t} = \frac{2}{c}\ \left| \boldsymbol{r}(t+\Delta t) - \boldsymbol{R}(t+\Delta t) \right|\ +\ 2\,\Delta t_{\mathrm{trop}} \ , \qquad (8.84)$$

where $\Delta t \approx \frac{1}{2}(\Delta t_1 + \Delta t_2) = \frac{1}{2}\,\Delta\tilde{t}$ is the propagation time between the position vectors $\boldsymbol{R}(t+\Delta t)$ and $\boldsymbol{r}(t+\Delta t)$.

One equation of type (8.84) is not sufficient to determine all parameters in the above list. Parameter determination tasks in satellite geodesy ask for a common processing of all observations of type (8.83) made by a global (or at least regional) network of tracking sites. Moreover, all observations made in a longer time-interval (a few days up to months or even years) have to be analyzed together in order to achieve a good separation of all parameters.

It is possible to define "restricted" problems. For permanent sites it is often allowed to assume the coordinates and the motion of the sites as known (e.g., for pure orbit determination). Two restricted problems, namely that of optimizing the measurement process in real time and that of screening the observations of a satellite pass are briefly addressed below.

Figure 8.11 shows the global network of the International Laser Ranging Service (ILRS) (see Table 2.5).

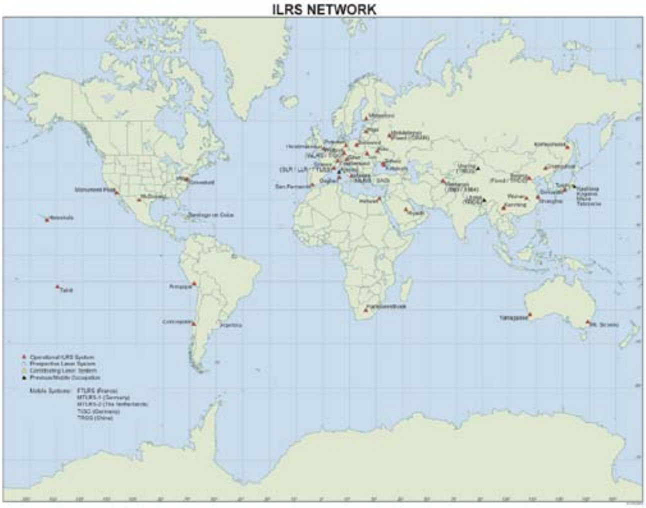


Fig. 8.11. The ILRS network

When compared to the global IGS network in Figure 2.7 the number of stations is much smaller (about 30 stations). The roots of the ILRS network are going back to the early 1970s, when the first sizeable SLR experiments took place. Seen from this perspective the ILRS network is much older than the IGS network. In order to facilitate collocation between technique-specific networks every ILRS observatory is (should be) equipped with a geodetic GPS receiver, making the ILRS network a subnetwork of the IGS network, as well.

As opposed to the IGS network where the equipment is rather small and homogeneous, the ILRS sites have a high degree of individuality. Figure 8.12 shows the ILRS part of the Zimmerwald geodynamics observatory. The observatory is equipped with an astronomical telescope of 1 m aperture and a Titanium-Sapphire Laser with a repetition rate of 10 Hz, a pulse width of 100 ps (corresponding to a length of 3 cm of the individual Laser pulse), and a wavelength of $\lambda = 423$ nm.

The Zimmerwald telescope was designed as a multipurpose instrument. The astrometric positions of the geostationary satellites analyzed in the previous sections of this chapter were made with this telescope, as well. The setup allows it to observe directions *and* distances in one and the same satellite pass with state-of-the-art accuracies.

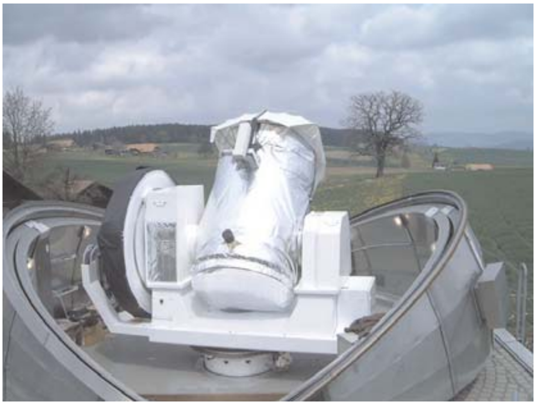


Fig. 8.12. The Zimmerwald Observatory

As outlined in section 2.2 the ILRS observatories measure the light travel times Δt of the Laser pulses from the observatory to the satellite and back to the observatory. As the field of view of a telescope of the type shown in Figure 8.12 is very small (about $20''$) and because the beam divergence of the Laser pulses is very small as well ($10''$) rather precise predictions of the satellites to be observed are required.

Figure 8.13 illustrates the signature of a typical Laser pass. When Lageos 1 (looking exactly like Lageos 2 in Figure 2.4) was first observed in this pass, the light travel times were about 53 ms, corresponding to a distance $\Delta \overset{\text{def}}{=} |\boldsymbol{r} - \boldsymbol{R}|$ of about 7950 km. Afterwards, the light travel times continually decreased to reach a minimum at the epoch of the closest approach (the minimum $\Delta t \approx 39$ ms corresponds to a distance $\Delta \approx 5850$ km). After the closest approach, the light travel times grow again. The entire pass lasted for about 37 minutes. A total of more than 9000 Laser pulses were sent out, a real-time analysis during the pass accepted about 1800 measured light travel times as candidate echoes. A screening procedure (briefly described below) eventually accepted about 900 measurements as real echoes. This performance is typical for day-light passes. The night-time performance (as judged from the ratio of accepted echoes and the total number of pulses) is about 10 times better. The gaps between the recorded data in Figure 8.13 are not due to instrument failures. They are caused by the fact that other Laser satellites in lower orbits were tracked during the time period of the Lageos 1 pass.

Figure 8.14 illustrates the parameter estimation process taking place during the satellite pass: The predicted orbit of the satellite and the (rather well) known geocentric coordinates of the observatory allow it to compute the difference between the predicted ranges and the actual measurements. Predictions are never 100% true – or do you believe in weather predictions?

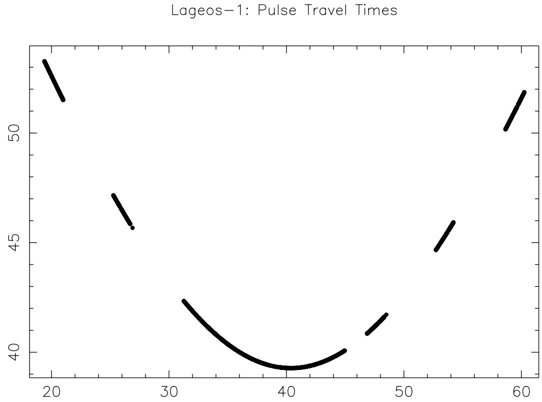


Fig. 8.13. Observed light-pulse travel times (ms) to Lageos 1, observed at Zimmerwald Observatory on July 7, 2002, $07^{\mathrm{h}}19^{\mathrm{m}} - 08^{\mathrm{h}}00^{\mathrm{m}}$

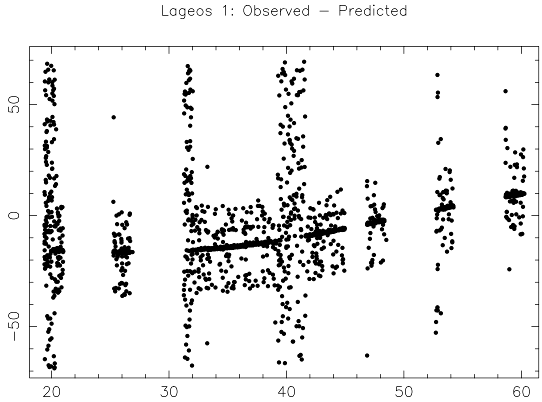


Fig. 8.14. Observed-predicted light travel times (ns) to Lageos 1, observed at Zimmerwald Observatory on July 7, 2002, $07^{\mathrm{h}}19^{\mathrm{m}} - 08^{\mathrm{h}}00^{\mathrm{m}}$ UT

In the case of a satellite orbit the uncertainty is mainly along track, i.e., all orbital elements may be assumed to be perfectly known, except one, the mean anomaly $\sigma_0 = \sigma(t_0)$ at the initial epoch t_0. The predictions allow it to define a so-called range gate. Only "echoes" within an observer-defined interval centered at the predicted ranges (originally ± 60 ns in Figure 8.14) are considered as candidate echoes. With a relatively high degree of probability it is then possible to decide in real time whether or not a registered light travel time within the range gate is real or not by looking for "identical" values in a list of recently established values "observed-predicted". If three or more of these coincidences are found (by a majority voting technique), the accepted observations may be used to determine an improved value for σ_0. From this time onwards, the predictions are much more reliable, allowing it to narrow down the range gate. Obviously, this was done in the example illustrated by Figure 8.14. After longer breaks, the range gate is reset to the original value. Unnecessary to say that the determination of σ_0 was based on the variational equation for the element σ_0 .

Real time decisions have to be based on a limited amount of information. This is why after a satellite pass a more correct analysis, this time based on a full orbit improvement process with all six elements (usually no dynamical or pseudo-stochastic parameters have to be introduced for the short arc of one pass), is performed before the data are sent to the ILRS data centers. As opposed to an orbit determination based on astrometric positions it is in general not possible to accurately determine all six osculating elements using the range observations of one observatory only, even if the individual observations are errorfree, because the normal to the orbital plane is very poorly defined (it may in essence rotate on a cone with the axis pointing from the geocenter to the observatory). In practice this means that the orbital elements i and Ω have to be slightly constrained in the analysis. The orbit improvement is done iteratively, where in principle the flag of every observation (indicating whether the observation will be used for the next step) may be redefined in each step. This process, the result of which is documented in Figure 8.15, is fully automatic. The rms error of this particular pass was 0.155 ns, corresponding to a mean error of about 2 cm in the measured ranges.

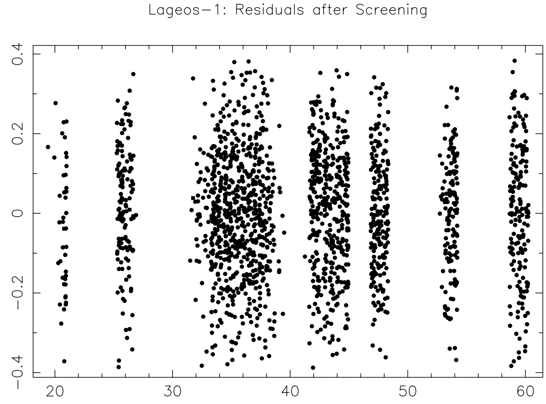


Fig. 8.15. Screened residuals (ns) of Lageos 1 observations at Zimmerwald Observatory on July 7, 2002, $07^{\mathrm{h}}19^{\mathrm{m}} - 08^{\mathrm{h}}00^{\mathrm{m}}$ UT

One of the primary targets of SLR, if not *the* primary target, is the determination of the Earth's stationary and (to some extent) the time-variable gravity field. The harmonic functions' coefficients C_{nm} and S_{nm} (see eqn. (3.150) and Table 3.1) are determined as dynamical parameters in SLR analyses combining many years of observations of all reasonable Laser satellites, where observations of Lageos 1, Lageos 2, and Starlette play a key role. These satellites, with their favourable area to mass ratios (see Table II- 3.4) were especially designed for gravity field determination. Only very few parameters have to be "sacrificed" to model non-gravitational effects like radiation pressure (for the Lageos satellites) and air drag (for Starlette). The parameter estimation procedures and the (historical) development of the two fundamental Earth gravity field models GEM and GRIM are described by Reigber in [89].

With every set of Earth potential coefficients it is possible to construct equipotential surfaces. The equipotential surface at mean sea level, the geoid, is of particular importance, e.g., for satellite altimetry. Figure 8.16 illustrates this equipotential surface w.r.t. a spheroid (where the height differences are "slightly" exaggerated).

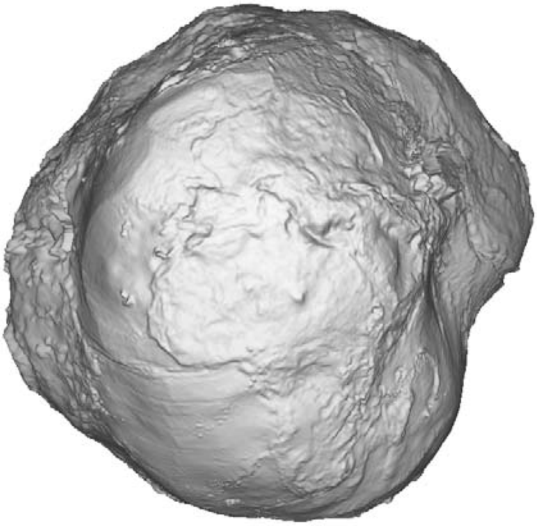


Fig. 8.16. The geoid with exaggerated vertical scale (Courtesy of Dr. Alain Geiger and Etienne Favey, ETH-Zürich, Switzerland)

Figure 8.16 symbolizes the historical contribution of SLR to geodesy and Earth sciences.

The station coordinates derived from SLR analyses are automatically referred to the geocenter, the Earth's center of mass (see discussion in section 3.4.2). Thanks to the insensitivity of the SLR satellites w.r.t. non-gravitational forces, geocenter determinations (i.e., determination of the center of mass w.r.t. the polyhedron of tracking sites) from SLR are accurate and reliable.

When the IERS was established in 1988, the SLR technique was (in essence) one of the two space geodetic techniques routinely determining the polar wobble coordinates x and y. SLR still contributes to that service today. Usually, one set of values is determined per three days. In principle it would be possible to estimate these parameters with a higher time resolution, but the observation density (due to weather and, to a lesser extent, day-time limitations) prevents this. With the advent of the GPS (see next section) this particular application of SLR lost some of its attractiveness. The importance of SLR thus resides in the determination of the gravity field and of the geocenter.

In addition, SLR is gaining more and more attention as a calibration tool for microwave observation techniques. SLR is in a position to play this role because its observable (8.83) is, as mentioned, only weakly affected by non-geometrical effects: tropospheric refraction may be accounted for using standard meteorological equipment, there are (in principle) no technique-specific biases. (The station-specific range biases usually have a rather high time stability, allowing it to keeping them constant over a long time). Figure 8.17, showing the differences w.r.t. GPS-derived orbits of all ranges measured by the ILRS to the GPS satellites equipped with SLR reflectors (PRN 5 and PRN 6) in the years 2001 and 2002, illustrates this calibration role. The constant bias of about − 5 cm, but also a periodic signal with an annual period are not yet fully understood. In view of the simplicity of the SLR observable these discrepancies must, however, probably be attributed to the GPS-derived orbits. Unnecessary to point out that such cross-technique checks are extremely important. Figure 8.17 is taken from [59].

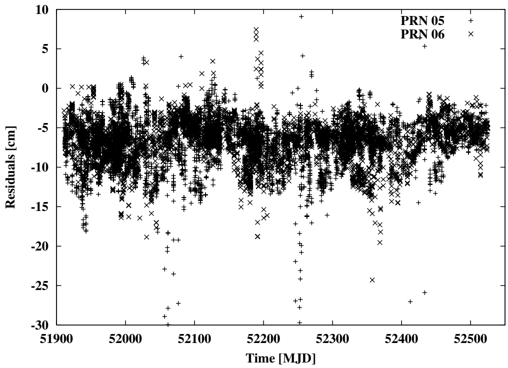


Fig. 8.17. Residuals of GPS satellites PRN 5 and 6

8.5.3 Scientific Use of the GPS

The *International GPS Service (IGS)* is an outstanding organization exploiting the US *Global Positioning System (GPS)*. Both, the GPS and the IGS, were already mentioned in Chapter 2. Subsequently, the GPS observables are defined in mathematical terms and the fascinating work of the IGS is highlighted with a few examples in this section.

The principle of the GPS code measurement is simple. The signal (or code) at an epoch characterized by an index l, emitted by a satellite j, contains the emission time τ_l^j of the satellite clock at signal transmission time and the receiver i keeps track of the time t_{il} of signal reception. The difference *reception time − emission time*, after multiplication with the speed of light c, is in essence the distance between the satellite and the receiver for a particular epoch. This idealized statement neither accounts for the propagation characteristics of the signal nor for the satellite and receiver clock errors relative to an ideal clock, the so-called GPS time. The real observable is referred to as the *pseudorange* p_{il}^j between satellite and receiver. It is the difference $c \cdot (\textit{reception time} - \textit{emission time})$, where the signal reception time is read from the receiver clock, and signal reception time from the satellite clock. Therefore the pseudorange may be decomposed as follows:

$$p_{il}^j = \rho_{il}^j - c\,\Delta t_l^j + c\,\Delta t_{il} + \Delta\rho_{il_{\text{ion}}}^j + \Delta\rho_{il_{\text{trop}}}^j + \epsilon_{il_{\text{cod}}}^j \,, \tag{8.85}$$

where

c is the speed of light,

p_{il}^j the pseudorange,

j the satellite index, i the receiver index, and l the epoch index;

$\rho_{il}^j \overset{\text{def}}{=} \left| \boldsymbol{r}\left(\tau_l^j\right) - \boldsymbol{R}(t_{il}) \right|$ is the geometric distance between the satellite at signal emission time τ_l^j and the receiver at signal reception time t_{il} (ρ_{il}^j also is referred to as *slant range* between satellite and receiver);

Δt_l^j is the error of the satellite clock w.r.t. GPS time at emission time,

Δt_{il} the error of the receiver clock w.r.t. GPS time at signal reception time;

$\Delta\rho_{il_{\text{trop}}}^j$ is the tropospheric range correction,

$\Delta\rho_{il_{\text{ion}}}^j$ is the ionospheric range correction, and

$\epsilon_{il_{\text{cod}}}^j$ is the measurement error of the observation.

The same signal (code) is sent on two different wavelengths $\lambda_1 \approx 19$ cm and $\lambda_2 \approx 24$ cm through the atmosphere. The corresponding carrier phases are also referred to as L_1 and L_2. Therefore, there are two equations of type (8.85) at each epoch (for each pair of satellite and receiver). Apart from the measurement error, only one term, namely the ionospheric refraction, is wavelength-dependent. This dependence is exploited for ionosphere modelling (see remarks at the end of this section).

When comparing the above observation equations with the corresponding equations (8.83, 8.84) of Laser distance measurements, one notes that two clock terms and the ionospheric refraction correction have to be taken into account in the case of GPS observations. These biases have to be dealt with when analyzing GPS data. The clock terms may only be determined accurately (or eliminated) from eqns. (8.85), if each receiver observes more than one satellite (quasi-simultaneously), and if the observations from more than one receiver are analyzed together. A new clock term has to be set up for each observation epoch.

There are two observations of type (8.85) per receiver and satellite, one for each of the two carrier frequency L_1 and L_2, which allow it to eliminate the ionospheric refraction term $\Delta\rho^j_{il_{\rm ion}}$ by forming the so-called *ionosphere-free linear combination* of these two equations (8.85). The procedure is possible only because the ionosphere is a dispersive medium: the ionospheric refraction correction is proportional to $\frac{E}{\nu^2}$, where E is the total electron content between receiver and satellite and $\nu = \frac{\lambda}{c}$ is the frequency of the signal. Alternatively, one may form the plain difference between the two equations for L_1 and L_2, which allows the direct determination of the ionospheric correction term. This latter procedure is, e.g., used to determine global models for the total electron content (discussed below).

Tropospheric refraction is delicate to model in the case of microwave observations because it is composed of the hydrostatic constituent (which also figures in the Laser observation equation) *and* the so-called *wet constituent*, which is due to the water vapor in the Earth's atmosphere, more precisely due to a resonance of water molecules with microwave radiation. The wet constituent is relatively small (a few percent compared to the hydrostatic part under "normal" (mid latitude) conditions, up to 10 – 25% in a tropical environment). Its high variability in space and time makes it extremely difficult to account for the wet delay with sufficient accuracy using standard meteorological measurements made at the receiver site. For high-accuracy applications of the GPS it is therefore mandatory to implement rather complex modelling procedures, which imply the estimation of many parameters for each receiver location, or, alternatively, to introduce the tropospheric refraction for each receiver as a stochastic quantity with given properties.

At observation time t_{il} not only the code, but also the phase ϕ^j_{il} may be exploited as an observable. The essential differences of the phase w.r.t. the code observable are

- the much better accuracy of the phase measurement (millimeters rather than (deci-)meters),
- a bias term $\lambda\, N^j_i$, the number N^j_i being an unknown integer, and
- the opposite sign of the ionospheric refraction term (the signal delay due to the ionosphere (in the case of the code observation) has to be replaced by the phase advance for phase observations).

As the receiver keeps track of the number of cycles received from the satellite-emitted signal as long as the satellite is above the receiver's horizon, there is only one such unknown parameter per receiver and satellite pass. If the signal is obstructed (or if the observation is disrupted for other reasons), a new so-called ambiguity parameter N_i^j has to be set up for the considered pair of receiver and satellite. With only minor neglects the observation equation for the GPS phase observable reads as

$$\phi_{il}^j = \rho_{il}^j - c\,\Delta t_l^j + c\,\Delta t_{il} - \Delta\rho_{il_{\text{ion}}}^j + \Delta\rho_{il_{\text{trop}}}^j + \lambda\,N_i^j + \epsilon_{il_\phi}^j \,, \qquad (8.86)$$

where N_i^j is the unknown *initial phase ambiguity parameter* or simply *ambiguity parameter*, λ is the wavelength of the carrier considered, and $\epsilon_{il_\phi}^j$ is the phase measurement error, which is of the order of a few mm only, for phase observations.

Exactly as in the case of code observations there are two observations of type (8.86) associated with the two carriers L_1 and L_2 . The two observations allow it to form the ionosphere-free observation (by eliminating ionospheric refraction) or the geometry-free linear combination, leaving only the corresponding linear combination of ionospheric refraction and of the ambiguity parameters in the equation.

Using the quasi-simultaneous phase observations of two satellites made from two receivers, one may form the so-called *double difference phase observation*, which is (almost) free of the clock errors, and where the remaining ambiguity term is known to be integer. Using statistical criteria, it is often possible to *resolve* the double difference ambiguities *after* the estimation process. The principles of ambiguity resolution are simple: If the real-valued estimates are close to integers and if the mean errors associated with them are small, the real valued estimates may be replaced by the integers, and the adjustment may be repeated by assuming N_i^j to be a known integer quantity in equation (8.86). The resolution of ambiguities in general leads to a much better determination of the remaining parameters – those of physical interest. Whereas the principles of ambiguity resolution are simple, the actual implementation is rather complex. The degree of difficulty increases with the size of the network and decreases with the length of observation time span. For more information concerning ambiguity resolution and other subtleties of scientific processing of the GPS observable the reader is referred to [122].

Let us now review some of the results generated by the IGS, which are based on the observations made by the IGS network (see Figure 2.7). The data are normally retrieved on a daily basis from the global IGS Data Centers and analyzed by the IGS Analysis Centers, which issue so-called *rapid* and *final* products. Currently there are eight IGS Analysis Centers.

Rapid IGS products are available with a delay of about one day, final products with a delay of about ten days. IGS products, which are provided for each day, include

- satellite orbits with an accuracy of about 5 cm,
- satellite clocks with an accuracy of about 0.05 ns and a time resolution of five minutes,
- daily values of polar motion components accurate to about 0.1 mas (milliarcseconds), corresponding to about 3mm on the Earth's surface,
- LOD (length of day) estimates with an accuracy of about 30 μs,
- tropospheric path delays for a selected number of stations with a time resolution of two hours, and
- station coordinates for those stations which are not known with sufficient (sub-cm) precision.

Initially, the IGS was designed as an orbit determination service for GPS satellites. Figure 8.18 documents this aspect of the IGS activities. It is based on the analyses performed every week by the IGS Analysis Coordinator, who compares and combines every week the ephemerides generated by the IGS Analysis Centers. The Analysis Center Coordinator, stemming from one of the IGS Analysis Centers, coordinates the work of the Analysis Centers. The IGS final orbits are the basis for most satellite-based national or international first order surveys since 1993.

Figure 8.18 shows the mean errors of the satellite positions (as produced by the IGS Analysis Centers) w.r.t. the combined orbit, which is established as a weighted average of the contributing centers. This rms error may be interpreted as a measure of consistency between the GPS ephemerides of the individual IGS Analysis Centers. It is a proud achievement of the IGS that this consistency level is today of the order of a few cm.

In the same process the satellite clocks and the Earth rotation parameters are analyzed and combined as well, with the goal to provide one consistent set of orbit, clock, and Earth rotation parameters, referring in particular to one consistent reference frame. The consistency of Earth rotation parameters is of the order of fractions of milliarcseconds, corresponding to daily values of the pole position of a few millimeters accuracy on the Earth's surface.

Some of the parameters routinely estimated by the IGS Analysis Centers might in principle be taken over as known from other sources (e.g., from VLBI analyses), which seem better suited for their determination. This is in particular true for the Earth orientation and rotation parameters (x and y coordinates of the Earth's rotation axis w.r.t. the Earth-fixed system and the length of day values). It became clear pretty soon after the beginning of the IGS Test Campaign in 1992, however, that no other technique was capable of providing this information with sufficient accuracy and resolution at the time needed by the IGS Analysis Centers.

This circumstance and the excellent global coverage provided by the IGS network (see Figure 2.7) are the primary reasons why IGS-derived polar motion

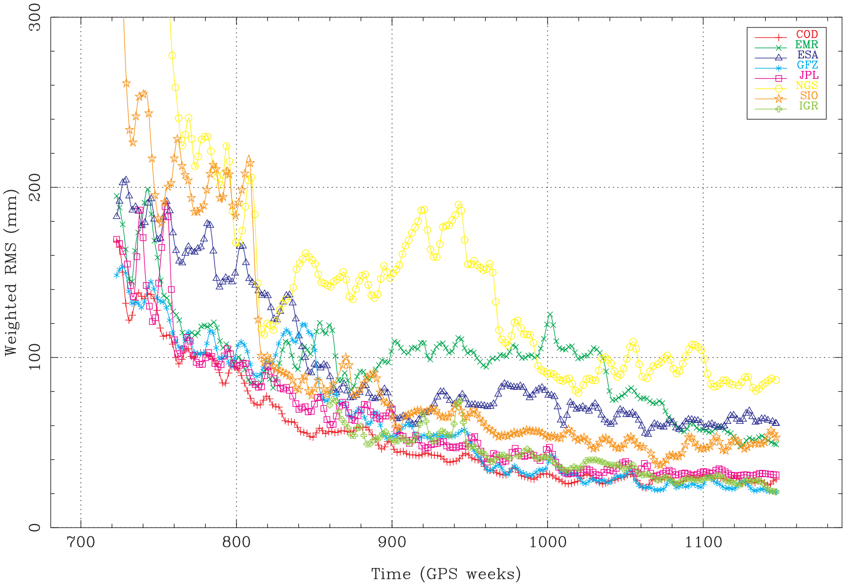


Fig. 8.18. Mean errors of the orbits as estimated by the IGS Analysis Center Coordinator between 1993 and January 2002

data are among the best available today. Since 21 June 1992 (the official start of the IGS Test Campaign) uninterrupted series of GPS orbits and of polar motion and length of day data are available through the IGS.

Figures 8.19 to 8.21 display polar motion as derived by the CODE (Center for Orbit Determinattion in Europe) Analysis Center of the IGS, which is located at AIUB (Astronomical Institute, University of Bern). It is a joint venture of AIUB, swisstopo (Swiss Federal office of Topography), the German BKG (Bundesamt für Kartographie und Geodäsie), and the French IGN (Institut Géographique National). Figure 8.20 is particularly instructive, because the time element (the third component of this three-dimensional representation) shows very nicely the beat period of about 6.2 years resulting from the (principal) Chandler period (named after its discoverer Seth Carlo Chandler (1846–1913)) of about 435 days and the annual period of 365.25 days (see Chapter II- 2). The results shown in Figures 8.19, 8.20 and 8.21 are in turn analyzed by the International Earth Rotation Service (IERS) together with those of all IGS, VLBI, and SLR Analysis Centers and used to produce the official IERS Earth rotation series. These IERS series were, e.g., used to compute the so-called angular momentum functions in Chapter II- 2.

In addition to the results already described, the IGS Analysis Centers perform each week coordinate solutions of their portion of the IGS network, which are then combined by the IGS Analysis Center Coordinator. Subsequently the IGS coordinate time series together with the contributions of the other

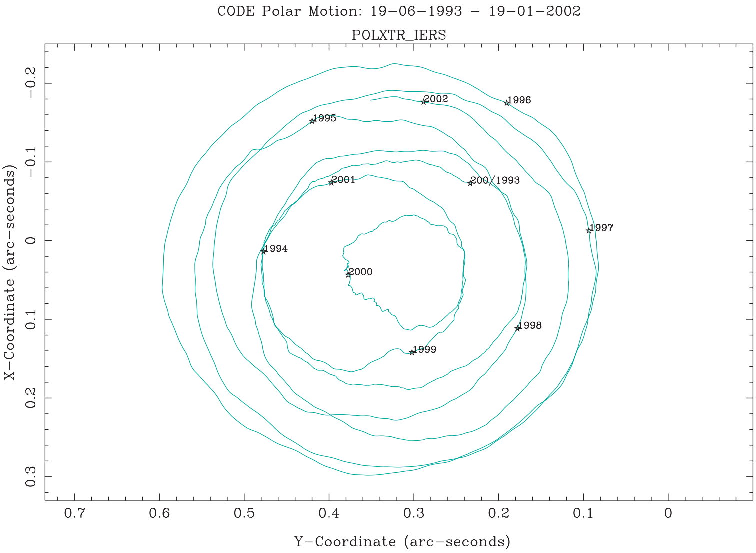


Fig. 8.19. Polar motion from CODE Analysis Center of the IGS (1993-2002)

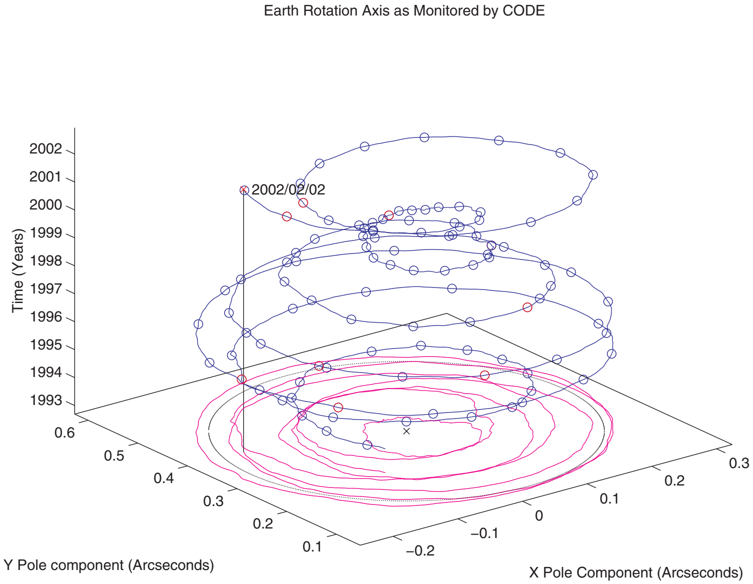


Fig. 8.20. Polar motion from the CODE Analysis Center of the IGS (1993-2002); vertical axis=time

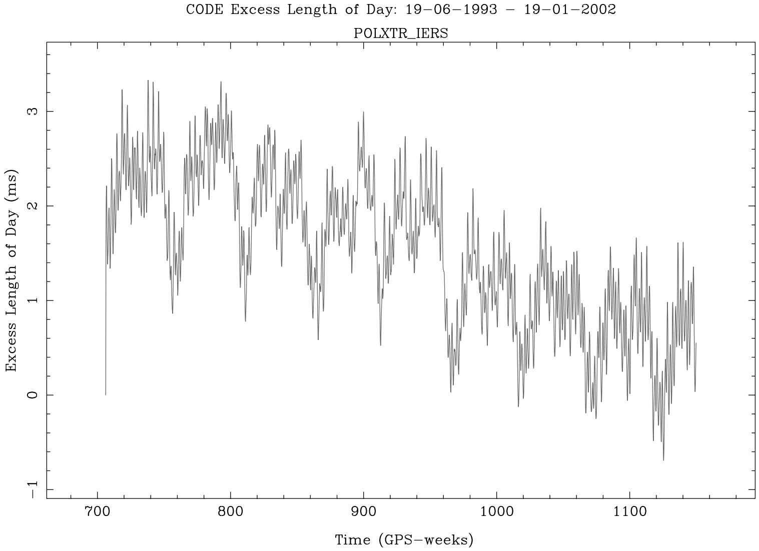


Fig. 8.21. Excess LOD from the CODE Analysis Center of the IGS (1993-2001)

space geodetic techniques are used by the IERS to establish the International Terrestrial Reference Frame (ITRF).

Not only the coordinates, but also the "velocities" of the observing sites relative to a Tissérand coordinate system may be extracted from the weekly coordinate estimates. Figure 8.22 illustrates the station velocities derived from the series of daily station positions produced by the CODE Analysis Center. Obviously it is nowadays possible to monitor plate tectonics so to speak "in real time" using space geodetic techniques. Remember that continental drift was but a postulate by Alfred Lothar Wegener (1880–1930) early in the 20th century!

More and more, the IGS network is used for purposes other than geodesy: The IGS network has, e.g., been enlarged to include time and frequency transfer. Thanks to the two carriers L_1 and L_2 it is possible to calculate the electron content along the line of sight between satellites and receivers for each individual observation. If one assumes that all electrons are contained in *one single layer* of "infinitesimal thickness" at a given height H above the spherical Earth, the electron content E measured along the line of sight receiver → satellite may be represented by a single layer density E_s at the height H:

$$E = \frac{E_s}{\cos z'} \ ,$$

where z' is the zenith distance of the line from the receiver to the satellite at the intersection point of this line with the single layer (the so-called iono-

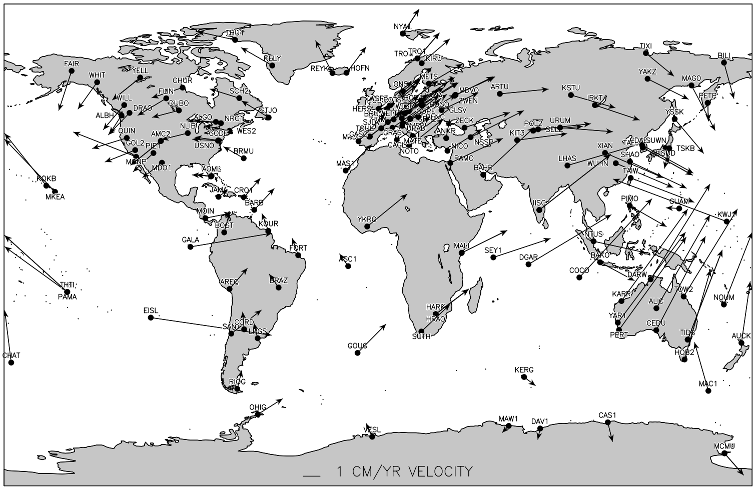


Fig. 8.22. Station motions of IGS sites from the CODE Analysis Center of the IGS

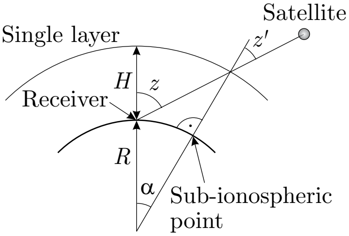


Fig. 8.23. Single layer model for the Earth's ionosphere

spheric pierce point). The formula is illustrated by Figure 8.23. By using the code and phase observations observed by the entire IGS network, the individual values E_s referring to a certain time interval may be used to model the vertical electron content E_s as a function of geographical (or geomagnetic) latitude and longitude $E_s = E_s(\phi, \lambda)$. Several Analysis Centers produce and make available ionosphere models based on the observations performed by the IGS network.

Figure 8.24 shows one such ionosphere map for February 5, 2002, 15^{h} UT. It fits the vertical electron contents stemming from about 100 receivers of the IGS network by a series of harmonic functions up to degree and order 15 in ϕ and λ. Two hours of data (between 14^{h} and 16^{h} UT) were used to generate this figure. Dark areas correspond to a high electron density, bright areas to a low electron density. The highest densities are observed near the sub-solar point (intersection of the line geocenter → Sun with the Earth's surface).

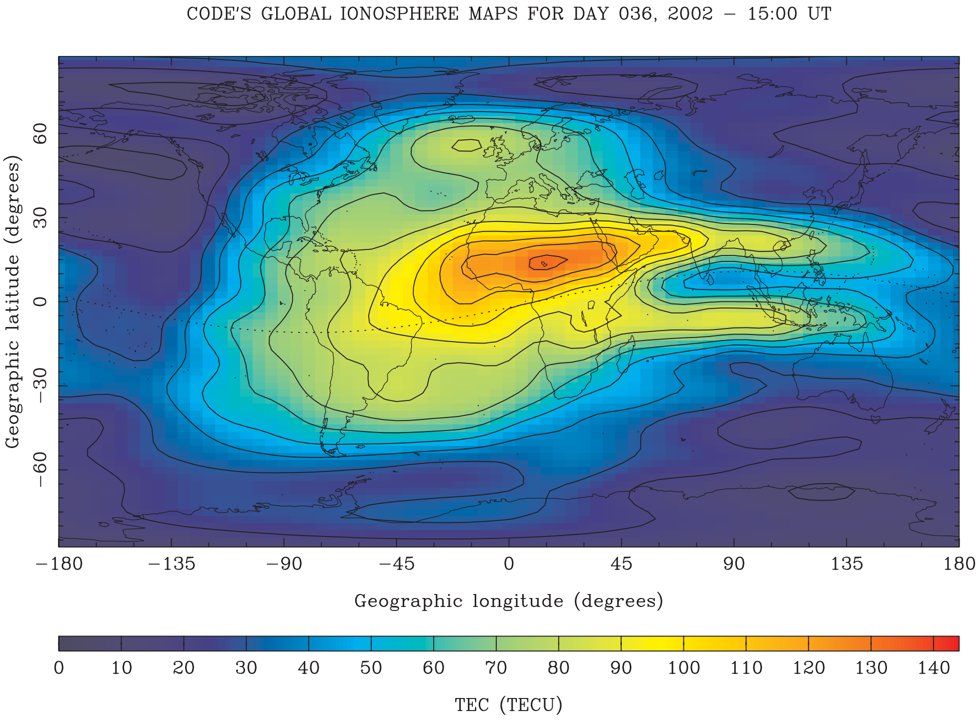

CODE'S GLOBAL IONOSPHERE MAPS FOR DAY 036, 2002 − 15:00 UT
Geographic latitude (degrees)
60
30
0
−30
−60
−180
−135
−90
−45
0
45
90
135
180
Geographic longitude (degrees)
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
TEC (TECU)

Fig. 8.24. Ionosphere map for February 5, 2002, 15^{h} UT

The equatorial bifurcation of the electron distribution is also clearly visible. Twelve maps like that provided in Figure 8.24 are produced every day. The figures and the corresponding coefficients of the development are available over the world wide web (http://www.aiub.unibe.ch/index.html).

The mean global total electron content (global TEC) is represented by the first term of the development. If this mean global TEC is drawn as a function of time one in essence monitors the effect of the solar activity on the Earth's upper atmosphere. Figure 8.25 shows the development of the mean TEC since January 1995. The Figure documents the effect of the maximum of solar activity in 2001 on the Earth's atmosphere. Variations with a period of about 27 days are due to the rotation of the Sun. Annual and semiannual periods are observed on top of these short period sisgnals. The smooth line (after January 2002) is a prediction based on the measured TEC values. For more information concerning the ionosphere we refer to [97] and [98], for more information concerning the IGS and its interdisciplinary significance we refer to [17] and to the homepage of the IGS (http://igscb.jpl.nasa.gov/).

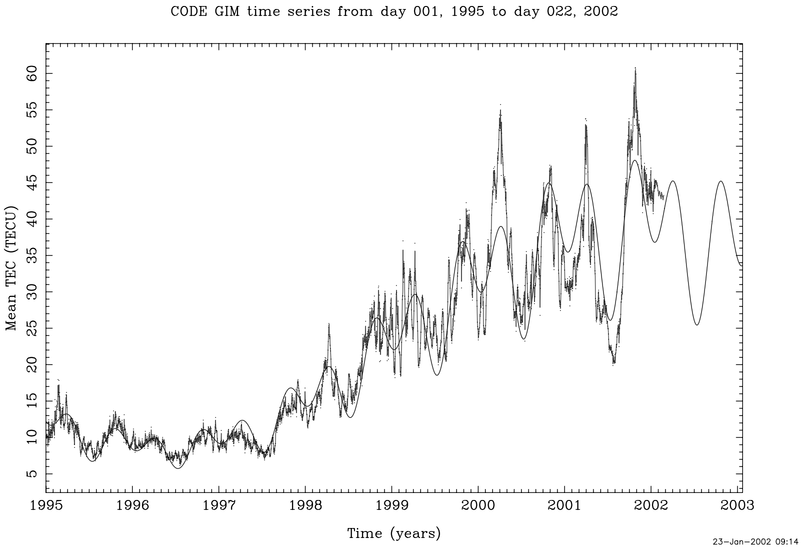


Fig. 8.25. Daily means of the total electron content (TEC) since 1 January 1995

8.5.4 Orbit Determination for Low Earth Orbiters

LEO Orbit determination based on the measurements of the onboard GPS receiver(s) is about to become a well-established, efficient, and robust technique. The accuracy requirements for LEO trajectories range between a few hundred meters to few centimeters or even millimeters. Naturally, the degree of difficulty grows with increasing accuracy requirements.

First dual-band spaceborne GPS receivers of good quality were already deployed in the early 1990s. Experience could be gained with data from receivers onboard TOPEX/Poseidon (see Figure 2.8), GPS/MET (GPS Meteorology using limb sounding), a satellite launched by JPL in 1995 to test the GPS occultation technique, etc. The first high-quality receiver well-suited for routine POD (Precise Orbit Determination) with 1 m accuracy or better is the so-called *blackjack receiver* developed by the JPL. Blackjack receivers are, e.g., used for the Argentine/US mission SAC-C, onboard the German/American research satellite CHAMP (CHAllenging Minisatellite Payload), and onboard the twin satellites GRACE A and GRACE B of the American/German GRACE mission.

Subsequently we will use a limited amount of data gathered by the GPS receiver onboard the CHAMP spacecraft. CHAMP, launched on July 15, 2000, is a combined mission to explore the Earth's gravity and magnetic fields. Figure 8.26 gives an impression of the spacecraft, Table 8.15 summarizes CHAMP's orbital characteristics (as of May 2001). CHAMP moves on an

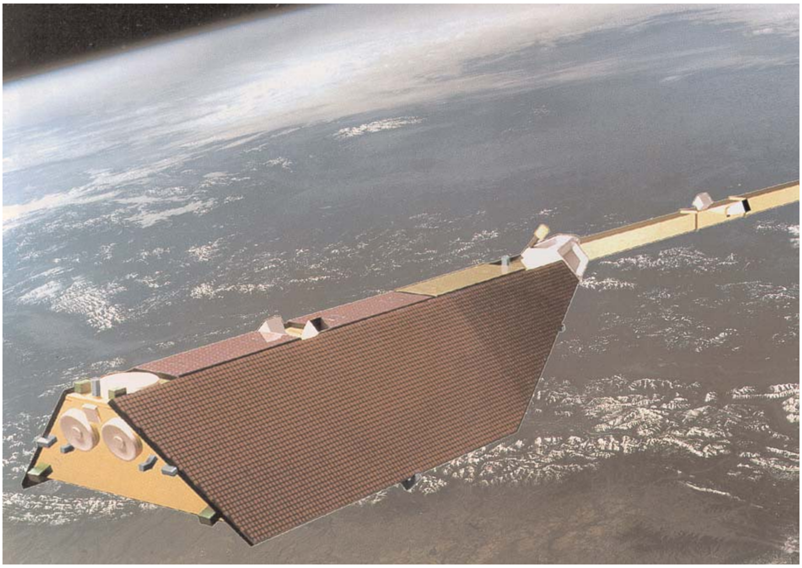


Fig. 8.26. The German satellite CHAMP

almost circular orbit in extremely low altitudes (about 430 km above the Earth's surface in May 2001). During its lifetime the orbital height gradually decreases. For precise orbit determination LEOs pose serious problems because of the insufficiently known force field, in particular air drag and higher-order gravity field parameters (which are at least initially unknown). The high inclination of CHAMP makes it an excellent probe mass for the determination of the Earth's gravity field. For more information concerning CHAMP we refer to the homepage of the GFZ (GeoForschungsZentrum) in Potsdam (http://op.gfz-potsdam.de/champ/index_CHAMP.html).

CHAMP has GPS receivers onboard for atmospheric sounding *and* for POD. Only the latter receiver and the antenna associated with it are of interest here. The zenith of the antenna of this POD receiver is pointing into the positive radial direction (seen from the geocenter). It is (barely) visible "on top of the roof" at the rear end (left) of the spacecraft in Figure 8.26.

Table 8.15. Osculating elements of CHAMP at t_0 = May 20, 2001, 0^{h} UT

Element	Value	Element	Value
a	6809 km	e	0.004
i	87.3°	Ω	34.6°
ω	178.5°	P	93 Min

The principles of LEO orbit determination are in essence the same as those underlying the determination of a trajectory of a moving (roving) receiver on the Earth's surface (on cars, bikes, pedestrians) or in the Earth's atmosphere (on airplanes): The position vector of the roving receiver (actually of its antenna) is determined at regular intervals (e.g., each second, every ten seconds) using the code and phase observations of all GPS satellites above the roving receiver's antenna horizon, recorded at the measurement epochs by the receiver. The measurement principle in its simplest form is illustrated by Figure 8.27. The code measurements of at least four GPS satellites observed at observation time t by the spaceborne GPS reciver are used to determine the geocentric position vector $\boldsymbol{r}(t)$ of the LEO's center of mass (neglecting the offset of the receiver's antenna w.r.t. the center of mass) and the clock offset of the spaceborne GPS receiver at time t, implying that four scalar unknowns have to be determined for each measurement epoch. The GPS satellites' state vectors $\boldsymbol{r}^j(t)$, $\dot{\boldsymbol{r}}^j(t)$ and the clock offsets of the GPS clocks w.r.t. GPS time are assumed to be known.

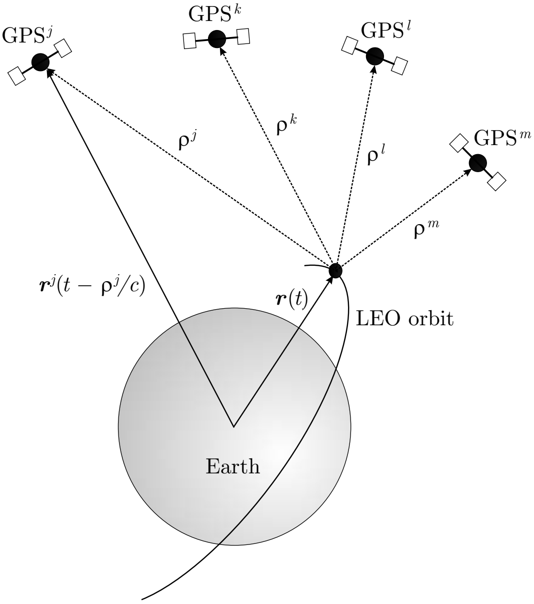


Fig. 8.27. LEO precise point positioning

When analyzing the observations stemming from spaceborne GPS receivers, one has to take into account the following characteristics differing from "normal" GPS data gathered on or near the Earth's surface:

- As the LEO is in essence in free fall in the Earth's gravity field, its trajectory may be well described as a particular solution of the equations of motion.

  This fact allows it to considerably reduce the number of unknowns: Instead of estimating one set of three coordinates for each geocentric LEO position vector $\boldsymbol{r}(t_l)$ at each measurement epoch t_l, one may simply solve for the orbit parameters, which does dramatically reduce the number of unknowns.
- Tropospheric refraction may be neglected for spaceborne GPS receivers.
- The "horizon" of the spaceborne GPS receiver is usually much better than the horizon of a roving receiver on the Earth surface, where the view may be obstructed by houses, trees, etc. Due to the height of the LEO above the Earth's surface it is in principle even possible to gather observations below the antenna's horizon. (Such observations are, however, often not of the best quality, because the antennas are not optimized for such observations).
- Whereas the GPS measurement scenario varies only slowly for receivers on the Earth's surface, the LEO GPS receiver sees a good part of the entire GPS constellation during one revolution (of about 1.5 to 2 hours). Rapidly changing measurement scenarios are the consequence.

LEO Orbit determination with spaceborne GPS receivers may be classified as follows:

1. **Classification according to the Orbit Model:**
   - *Kinematic methods* do *not* make use of the fact that the geocentric position vector $\boldsymbol{r}(t)$ of the spacecraft's center of mass is (in good approximation) the solution of the equations of motion. They introduce a new, unknown position vector for each measurement epoch of the onboard GPS receiver. The result of the orbit determination process thus consists of a table of satellite positions, the table's spacing being defined by the measurement rate of the onboard GPS receiver.
   - *Dynamic methods* model the geocentric position vector $\boldsymbol{r}(t)$ of the satellite's center of mass as a particular solution of the equations of motion. The unknowns of the parameter estimation process are, e.g., the initial osculating elements of an arc plus a wide variety of dynamical parameters (e.g., gravity field parameters, parameters associated with drag and radiation pressure, deterministic empirical parameters, etc.). The main advantage of dynamical methods w.r.t. kinematic methods resides in the greatly reduced number of orbit parameters (typically a few tens to hundreds as opposed to four unknowns per measurement

epoch when using kinematic methods). Dynamic methods require the knowledge of an a priori trajectory, solving the equations od motion.

- *Reduced Dynamics methods* lie logically between purely kinematic method and purely dynamic methods. The LEO trajectory is allowed to have a stochastic component, either by introducing pseudo-stochastic pulses (to be discussed below), or by introducing system noise into the equations of motion, which actually replaces the differential equations by stochastic differential equations. Many different parameterizations are possible. A wide variety of methods exists. Reduced dynamics methods of course also require the knowledge of an priori trajectory.

2. **Classification according to the GPS Analysis Strategy:**

- *PPP (Precise Point Positioning) methods* make use of eqns. (8.85) for the GPS code and eqns. (8.86) for the GPS phase observables *without applying any data differencing techniques.* This processing mode is also referred to as *zero difference GPS processing.* In its purest form, the PPP method takes over the GPS satellites' clock terms $c\,\Delta t_l^j$ and the orbital positions of the GPS satellites as known quantities from independent orbit determination procedures for the GPS satellites. Typically, the IGS products are used for this purpose. If the ionosphere-free linear combination of the original observations is analyzed, the observation equations (8.85) for code may be reduced considerably:

$$\tilde{p}_l^j = \rho_l^j \, + \, c\,\Delta t_l \, + \, \epsilon_{l_{\mathrm{cod}}}^j \ , \tag{8.87}$$

where the GPS satellite clock term was absorbed into the modified pseudorange term $\tilde{p}_l^j$.

Under these assumptions there are only four unknowns left in the equations (8.87) for each measurement epoch, namely the receiver clock term and, e.g., the three Cartesian coordinates in the inertial system of the spacecraft's position vector $\boldsymbol{r}(t_l)$. If a dynamical method is used, each $\boldsymbol{r}(t_l)$ has to be treated as a function of the orbit parameters, i.e., $\boldsymbol{r}(t_l) \overset{\mathrm{def}}{=} \boldsymbol{r}(t_l; p_1, p_2, \ldots, p_n)$. For purely kinematic procedures the code observations referring to different epochs are, however, mathematically independent and may be treated separately.

The phase observation equations (8.86) referring to the ionosphere-free linear combination may be reduced in a similar way as those for the code observations:

$$\tilde{\phi}_l^j = \rho_l^j \, + \, c\,\Delta t_l \, + \, \xi^j \, + \, \epsilon_{l_\phi}^j \ , \tag{8.88}$$

where $\xi^j \overset{\mathrm{def}}{=} \frac{\nu_1^2}{\nu_1^2 - \nu_2^2}\,\lambda_1\,N_1^j \, - \, \frac{\nu_2^2}{\nu_1^2 - \nu_2^2}\,\lambda_2\,N_2^j$ is the ambiguity term in the ionosphere-free linear combination, which may assume any real value. The GPS satellite clock terms are considered known (as in the case of the code observations). Orbit determination procedures making use of the phase observable are more elaborate, because of the ambiguity term ξ^j : It is no longer possible to treat each epoch independently (as it is easily possible with the code equations (8.87)).

The JPL (Jet Propulsion Laboratory) pioneered the point positioning approach [134]. Thanks to its simplicity and thanks to the optimum use made of the IGS products, the method is widely used today. The method is only limited by the accuracy of the GPS orbits and the GPS satellite clocks.

- *Double Difference methods* combine the original observations (code and/or phase) of the spaceborne receiver with those of Earth-fixed receivers. The method treats the satellite as a roving (kinematic) receiver in a global network of Earth-fixed stations. The associated parameter estimation process is based on the well-known double difference processing of kinematic GPS observations (see, e.g., [122] or [58]). Ambiguity resolution and other techniques related to the double difference processing may be applied in a straightforward way. The method is CPU intensive, but promises highest accuracy. Rothacher and Švehla from the Technical University of Munich, see, e.g., [115] and [116], are protagonists of this approach. They use a modified version of the Bernese GPS Software [58] for this approach. The observations from a world-wide network of receivers have to be used. The advantage of the method resides in the fact that the accuracy requirements regarding GPS orbits and clocks are greatly reduced.

Advanced Orbit Modelling: Pseudo-Stochastic Parameters. With only few exceptions the attitude of scientific satellites is actively maintained. Also, it may be necessary from time to time to perform small orbital manoeuvres to optimize the orbital characteristics of a satellite (an example was given in Chapter II- 3, Figure II- 3.15). When determining the orbit of a LEO like CHAMP one has moreover the problem, that the forces acting on the satellite are not fully known. It is, e.g., not possible to model air drag with the accuracy required. Also, the gravity field is (at least initially) not known with sufficient accuracy to allow for the determination orbital arcs of, let us say, one day or longer.

One radical method of curing orbit modelling deficiencies of this kind is to break up the original arc into shorter arcs. To a great extent the modelling deficiencies are then absorbed by the initial state vectors of the shorter arcs. This well known method usually is referred to as the *short arc method.* One should be aware of the fact, however, that this simple method multiplies the number of arc-specific parameters by the number of arcs generated – which may considerably weaken the solutions. Also, an old Latin proverb says *natura non facit saltus.*

Another method, very powerful and widely used technique to cope with this problem is to replace the deterministic differential equation systems describing orbital motion by *stochastic differential equation systems*, which contain on top of all deterministic forces so-called stochastic accelerations, which are characterized by the (known) mean values (usually the zero vector) and the associated (known) variance-covariance matrix. Stochastic modelling of this kind is possible without major problems, provided the classical least-squares approach is replaced by digital filter methods, in particular Kalman- and Kalman-Bucy filters. In the framework of this more general theory every deterministic parameter estimated may be replaced by a stochastic process. Not only the measurement noise, but also the system noise (represented by the stochastic constituent in the differential equations in the case of orbit parameters) has to be considered.

A general approach of this type is out of the scope of our treatment of orbit determination. We will instead introduce a method which allows for stochastic changes of the orbit, which may be established by conventional least-squares methods. The method is in essence equivalent to the method of Kalman filtering. The difference resides in the fact that the stochastic component is introduced on the level of the first, and not of the second derivative of the orbit. Let us mention that the pseudo-stochastic method was generalized/modified to allow for piecewise constant or even piecewise linear accelerations [61].

Our original method of pseudo-stochastic pulses allows for instantaneous velocity changes δv in predefined directions at predefined epochs t_i. An instantaneous velocity change δv at the epoch t_i in a predetermined direction $\boldsymbol{e}$ is called a *pseudo-stochastic pulse*. Depending on the application many such pulses (velocity changes) may be set up. Up to three pulses in different, not necessarily orthogonal directions may be set up at one and the same epoch t_i (it does not make sense to set up more than three pulses, because in three-dimensional space four or more vectors are always linearly dependent). Usually, pseudo-stochastic pulses are set up either in the radial, the along-track, the out-of-plane directions, or in a combination thereof. The spacing of the "stochastic epochs" t_i may be different for different components. It is even possible to define the spacing in a more general way (e.g., through a user-provided table of the epochs t_i). The velocity changes may be constrained to "reasonably small" values by introducing artificial observations of the velocity changes (details are provided below). The method of pseudo-stochastic pulses was introduced in [12]. It may be characterized as follows:

- Each orbital arc is continuous.
- Each arc is represented piecewise by conventional ordinary differential equation systems (deterministic equations of motion).
- At predetermined epochs t_i (e.g., every five minutes) the satellite is allowed to change its velocity instantaneously in (up to three) predetermined directions.

- Pseudo-stochastic pulses δv are in every respect "normal" parameters of a classical least-squares adjustment process.
- Each pseudo-stochastic pulse is characterized by an expectation value and a variance (to be more precisely defined below).

There is a certain degree of arbitrariness in our approach (exactly as in the case of the Kalman-Bucy filtering): The number of epochs and the associated variances may be selected in many different ways. On the positive side one may state that the procedure is very flexible: depending on the actual number and the statistical properties of the pulses the orbit may be either purely deterministic, purely kinematic, or something in-between, usually referred to as *reduced dynamics orbit.*

The sizes of the velocity changes are controlled by artificial observation equations

$$\delta v = 0 \ , \tag{8.89}$$

associated with the *prescribed weights*

$$w \overset{\text{def}}{=} \frac{\sigma_0^2}{\sigma^2(\delta v)} \ . \tag{8.90}$$

The scalar velocity change δv is thus constrained as a random variable with expectation value zero and variance $\sigma^2(\delta v)$. σ_0 is the mean error of unit weight of the adjustment.

If $\sigma(\delta v)$ is big, the weight w is small, which allows δv to assume rather big values. If $\sigma(\delta v)$ is small, only minor velocity changes are possible. The allowed velocity changes roughly lie within the range $\pm 3\,\sigma(\delta v)$.

In order to introduce the parameter δv into the adjustment process we need to know the partial derivative of the orbit $\boldsymbol{r}(t)$ w.r.t. this parameter. This derivative w.r.t. a particular pseudo-stochastic pulse δv having occurred at epoch t_i may be represented as a linear combination of the partial derivatives w.r.t. the six osculating orbital elements $I_k\,,\ k = 1, 2, \ldots, 6$ at the initial epoch t_0 . In order to prove this statement we assume that the velocity change δv refers to the direction represented by the unit vector $\boldsymbol{e}$ (e.g., one of the unit vectors of the $\mathcal{R}$-system defined in Table 4.3). The associated changes in the initial conditions of the LEO at time t_i may thus be written as:

$$\begin{aligned} \delta\dot{\boldsymbol{r}}(t_i) &= \delta v\, \boldsymbol{e} \\ \delta\boldsymbol{r}(t_i) &= \boldsymbol{0} \ . \end{aligned} \tag{8.91}$$

Due to this pulse the orbit will be modified for times $t \geq t_i$ according to (neglecting higher-order terms in δv):

$$\delta\boldsymbol{r}(t) = \left(\frac{\partial \boldsymbol{r}}{\partial\,(\delta v)} \right)(t)\ \delta v \ , \tag{8.92}$$

where

$$\begin{aligned}\left(\frac{\partial \boldsymbol{r}}{\partial\left(\delta v\right)}\right)(t_i) &= \boldsymbol{0}\\ \left(\frac{\partial \dot{\boldsymbol{r}}}{\partial\left(\delta v\right)}\right)(t_i) &= \boldsymbol{e}\ .\end{aligned} \tag{8.93}$$

The above equations may be interpreted as the initial values referring to epoch t_i of the homogeneous variational equations associated with the (deterministic) equations of motion. The variational equations associated with the equations of motion were treated in Chapter 5.

The variational equations associated with the equations of motion together with the above initial conditions define a particular solution of these variational equations. It is well-known, on the other hand, that each particular solution of a linear homogeneous system of differential equations may be written as a linear combination of a complete system of solutions. The six solutions of the variational equations corresponding to the six osculating elements $I_k\ ,\ k = 1, 2, \ldots, 6$ at the initial epoch t_0 form such a complete system. We may thus write the partial derivative w.r.t. the parameter δv as follows:

$$\left(\frac{\partial \boldsymbol{r}}{\partial\left(\delta v\right)}\right)(t) = \sum_{k=1}^{6} \beta_k \left(\frac{\partial \boldsymbol{r}}{\partial I_k}\right)(t) \overset{\text{def}}{=} \sum_{k=1}^{6} \beta_k\, \boldsymbol{z}_k(t)\ , \tag{8.94}$$

where the symbol I_k stands for one of the six osculating elements referring to the initial epoch t_0. The functions $\boldsymbol{z}_k(t)$ are the (known) solutions of the variational equations for the initial osculating elements I_k. The six coefficients $\beta_k\ ,\ k = 1, 2, \ldots, 6$, of the above linear combination are determined by the six condition equations (following from the initial conditions (8.93))

$$\begin{aligned}\sum_{k=1}^{6} \beta_k\, \boldsymbol{z}_k(t_i) &= \boldsymbol{0}\\ \sum_{k=1}^{6} \beta_k\, \dot{\boldsymbol{z}}_k(t_i) &= \boldsymbol{e}\ .\end{aligned} \tag{8.95}$$

Observe that the coefficients β_k are time-*in*dependent and may therefore be determined once and for all by solving one system of six linear algebraic equations for each pulse set up in the analysis. The workload associated with the computation of the partial derivatives w.r.t. the stochastic parameters is therefore (almost) negligible, even if their number is large.

When setting up hundreds of these pseudo-stochastic pulses, the dimension of the normal equation system grows considerably, which may lead to a very inefficient processing scheme, when following the standard least squares procedures. One may show, however, that the complete normal equation system may be set up as a function of the reduced normal equation system (containing only the deterministic parameters) – provided the contributions to the reduced normal equation are stored for all stochastic epochs t_i . This is possible because eqn. (8.94) says that each partial derivative w.r.t. a stochastic parameter may be written as a linear combination of the partial derivatives w.r.t. the initial osculating elements. This efficient procedure is implemented in program SATORB.

Simplified, Efficient and yet Precise LEO-POD. When striving for the highest possible accuracy, one of course has to use one of the (not so easily available) advanced software packages capable of optimally coping with this problem. They are based on one or several of the orbit determination procedures principles discussed previously. The Bernese GPS Software [58] is one example.

The two programs LEOKIN and SATORB of the attached program system (see Part III of this work) may be used to calculate precise LEO orbits. Purely kinematic, purely dynamic, and reduced dynamics orbits may be established. These programs are documented in Chapter II- 7. The principles underlying the two programs are rather simple: LEOKIN calculates a table of LEO positions and/or position differences (details provided below) using the kinematic method, program SATORB uses theses positions (and possibly position differences) as observations to determine a dynamical or reduced dynamics orbit. A conventional least squares adjustment procedure (making, however, use of the efficiency tools described in the previous paragraph) is underlied in program SATORB.

This simple procedure, using kinematically established positions and position differences (by LEOKIN) as intermediary observations (in SATORB), is not the best possible from the point of view of adjustment theory. It is of course preferable to use directly the GPS observations to determine the best possible (reduced dynamics) orbit. The approximate procedure is, on the other hand, very instructive, remarkably robust, and the achievable accuracies are sufficient for many (if not most) applications. The achievable accuracies are of the order of $0.5 - 1$ m (rms per coordinate) if only code observations are used, of the order of $1 - 2$ dm (rms per coordinate), if the phase observations are used, as well.

It was mentioned above that kinematic POD becomes much more elaborate as soon as the GPS phase observations are used in addition to the GPS code observations (because it is no longer possible to calculate the positions epoch by epoch). This statement is, as a matter of fact, only true, if the phase observations are used in the statistically correct sense. They may, however, be used in an alternative way, which still makes good (but not optimal) use of the accuracy of the phase observations:

- Instead of analyzing the phase observation eqns. (8.88) for each epoch, one uses the differences of such observations referring to subsequent epochs (pertaining to one and the same satellite). These differences are unambiguous.

- Instead of solving for the position vector $\boldsymbol{r}(t)$ at the observation epochs, one solves for the position difference vector of the satellite between subsequent epochs.

The "incorrectness" of the method in essence resides in neglecting the mathematical correlations between subsequent phase difference observations.

Let us now explain in detail how position differences may be determined from subsequent phase observation equations of type (8.88). It is essential that the difference of two such equations referring to one particular satellite (the satellite index j can be skipped, because the method may be explained using only one satellite) and the two epochs t_l and t_{l+1} does no longer contain the ambiguity term ξ:

$$\Delta\tilde{\phi}_l \overset{\text{def}}{=} \tilde{\phi}_{l+1} - \tilde{\phi}_l = \rho_{l+1} - \rho_l + c\left(\Delta t_{l+1} - \Delta t_l\right) + \epsilon_{l_\phi} \,. \tag{8.96}$$

The above equation plays the equivalent role for the determination of the position difference vector

$$\delta\boldsymbol{r}_{l,l+1} \overset{\text{def}}{=} \boldsymbol{r}_{l+1} - \boldsymbol{r}_l \tag{8.97}$$

as the equations (8.87) do for the determination of satellite position vectors $\boldsymbol{r}_l$. Figure 8.28 illustrates how the difference vector $\boldsymbol{r}_{l,l+1}$ is formed and how the distances $\rho_{\dots}$ have to be interpreted.

We will now assume that rather precise LEO position vectors are already available from the analysis of GPS code observations. We may thus assume to know the ranges ρ_l and ρ_{l+1} with an accuracy of a few (deci)meters. For further improvement we may linearize the ranges and range differences as a function of the unknown coordinate differences w.r.t. the known a priori positions:

$$\begin{array}{lll}
\rho_{l+1} & = (\rho_{l+1})_0 + \displaystyle\sum_{k=1}^{3}\left(\frac{\partial\rho_{l+1}}{\partial\rho_{l+1,k}}\right)_0 \Delta\rho_{l+1,k} & = (\rho_{l+1})_0 - \boldsymbol{e}_{l+1}\cdot\Delta\boldsymbol{r}_{l+1} \\
\rho_l & = (\rho_l)_0 + \displaystyle\sum_{k=1}^{3}\left(\frac{\partial\rho_l}{\partial\rho_{l,k}}\right)_0 \Delta\rho_{l,k} & = (\rho_l)_0 - \boldsymbol{e}_l\cdot\Delta\boldsymbol{r}_l \\
\hline
\Delta\rho_{l+1,l} & \overset{\text{def}}{=} \rho_{l+1} - \rho_l & = (\Delta\rho_{l+1,l})_0 \\
 & & \quad - \left[\,\boldsymbol{e}_{l+1}\cdot\Delta\boldsymbol{r}_{l+1} - \boldsymbol{e}_l\cdot\Delta\boldsymbol{r}_l\,\right] \\
\hline
\end{array} \tag{8.98}$$

The subscript "0" stands for the known a priori values, calculated with a known approximate orbit, $\boldsymbol{e}_l$ is the unit vector pointing from the LEO to the GPS satellite at time t_l (actually from the LEO at signal reception time to the GPS satellite at signal emission time).

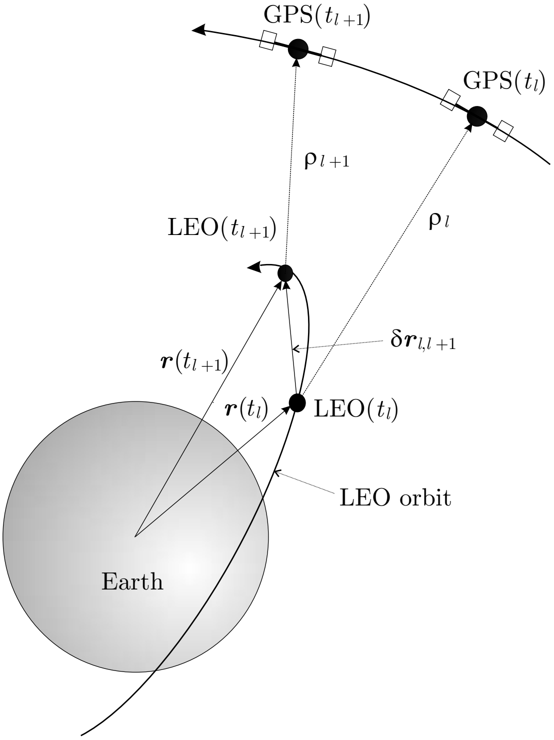


Fig. 8.28. Precise determination of LEO position differences $\delta \boldsymbol{r}_{l,l+1} \overset{\text{def}}{=} \boldsymbol{r}_{l+1} - \boldsymbol{r}_l$

We may now use eqn. (8.97) to replace the correction $\Delta \boldsymbol{r}_{l+1}$ of the position vector at epoch t_{l+1} by the correction of the position vector at t_l and the correction of the difference vector $\delta \boldsymbol{r}_{l,l+1}$. The third of eqns. (8.98) then reads as

$$\delta\rho_{l+1,l} = (\delta\rho_{l+1,l})_0 - \left[\boldsymbol{e}_{l+1} \cdot \{ \Delta \boldsymbol{r}_l + \Delta\delta \boldsymbol{r}_{l,l+1} \} - \boldsymbol{e}_l \cdot \Delta \boldsymbol{r}_l \right] , \qquad (8.99)$$

where

$$\Delta\delta \boldsymbol{r}_{l,l+1} \overset{\text{def}}{=} \delta \boldsymbol{r}_{l,l+1} - (\delta \boldsymbol{r}_{l,l+1})_0 \qquad (8.100)$$

is the correction of the a priori difference vector of subsequent LEO position vectors.

With a suitable rearrangement of terms the above equation may be brought into the form

$$\delta\rho_{l+1,l} = (\delta\rho_{l+1,l})_0 - [\boldsymbol{e}_{l+1} - \boldsymbol{e}_l] \cdot \Delta \boldsymbol{r}_l - \boldsymbol{e}_{l+1} \cdot \Delta\delta \boldsymbol{r}_{l,l+1} . \qquad (8.101)$$

The second term on the right hand side is a product of two small quantities which shall be neglected from now on:

$$\delta\rho_{l+1,l} \overset{\text{def}}{=} (\delta\rho_{l+1,l})_0 - \boldsymbol{e}_{l+1} \cdot \Delta\delta\boldsymbol{r}_{l,l+1} \ . \tag{8.102}$$

Clearly, this approximation introduces an error of the order of $10^{-3} - 10^{-4}$ of the neglected quantity $\Delta\boldsymbol{r}_l$. Assuming that the a priori position vectors were determined with accuracies of few (deci)meters, the error introduced in eqn. (8.102) is at maximum of the order of a few millimeters. As the errors of the point positioning with code are (better: should be) random, we thus accept random errors of the order of few millimeters by accepting the approximation (8.102).

If we use the approximation (8.102) in the observation equation for the phase differences (8.96), we obtain equations which allow us to estimate the position difference vector $\delta\boldsymbol{r}_{l,l+1}$ of subsequent satellite positions using the phase observations essentially with the accuracy dictated by the noise of the phase observable. The position difference vectors may thus be established with an accuracy of few cm.

The method is most attractive from the efficiency point of view. From the structural point of view it may be viewed as the equivalent to the point positioning method with the code observable. The observation equations for all available satellites corresponding to a particular epoch difference $t_{l+1} - t_l$ may be processed independently from those corresponding to other epoch differences, *provided* the mathematical correlations between the phase differences of subsequent epoch differences are gracefully ignored. For more information concerning this topic we refer to [24].

A Case Study: The Orbit of CHAMP. In order to illustrate the above theoretical developments, the observations made on May 5, 2001, by the CHAMP blackjack receiver, are processed in six different ways described by Table 8.16. The data used were made available by the GFZ for the LEO Working Group of the IGS. The permission to use these data is gratefully acknowledged.

Table 8.16. Orbital analyses performed with CHAMP data of May 5, 2001 (day 140 of year 2000)

Data	Program	n_o	n_det	n_s	rms w.r.t. R-Orbit [cm]	Result
Code	LEOKIN				402	C1-Orbit
Code	SATORB	8024	15	45	62	C2-Orbit
Code	LEOKIN				134	C3-Orbit
Code	SATORB	8663	15	429	48	C4-Orbit
Code & Phase	LEOKIN				14	P1-Orbit
Code & Phase	SATORB	16715	15	429	14	P2-Orbit

The first column of Table 8.16 says that the first four orbits are uniquely based on GPS code observations (see eqns. (8.87) and Figure 8.27), the last two on code and phase observations (see eqns. (8.102), (8.96) and Figure 8.28)). The second column identifies the program generating the resulting orbit (name in last column). The third column contains the number n_{o} of pseudo-observations, the fourth the number n_{det} of deterministic parameters (incl. the six osculating elements), and the fifth the number n_{s} of pseudo-stochastic parameters. The total number n_{par} of parameters thus is $n_{\mathrm{par}} = n_{\mathrm{det}} + n_{\mathrm{s}}$. Observe that LEOKIN is only capable of producing kinematic orbits, whereas SATORB generates (reduced) dynamic orbits using the positions (and possibly position differences) of the preceding LEOKIN orbit (in Table 8.16) as "observations". The total number of observations (actually used), the number of deterministic orbit parameters, and the number of stochastic pulses are provided in columns 3 – 5. Column 6 of Table 8.16 compares the resulting orbits with a reference orbit. The method of orbit comparison is that of a simplified Helmert transformation (similarity transformation), with only three translation parameters adjusted.

R-Orbit: The reference orbit, referred to as "R-Orbit" from now on, was produced at the TUM (Technical University of Munich) (see [115]) by processing all code and phase observations in one program run using the Bernese software package (see [58]). The R-Orbit deserves it to be used as reference: the phase and code observation equations (8.87) and (8.88) of the entire day were processed (with appropriate weighting) together in one program run. All parameters, i.e., the LEO orbit parameters (initial osculating elements, dynamical parameters, and pseudo-stochastic pulses), one clock term per observation epoch, and all initial phase ambiguity parameters, were simultaneously solved for. The number of orbit and clock parameters is considerable (measured in thousands). A few hundred ambiguity parameters had to be set up because of the rapidly changing observation geometry. The best possible gravity field available was used; a set of three pseudo-stochastic pulses was set up every six minutes.

Let us now consider the six orbits of Table 8.16 in more detail:

C1-Orbit: In a first step we assume that there is *no* a priori orbit information available. Program LEOKIN then produces a table of LEO positions using the GPS code observation equations (8.87), which are spaced by 10 seconds in our test data set. For each observation epoch all available code observations (8.87) are used to produce a PPP-solution. The GPS satellite clock corrections, assumed as known, stemmed from a special analysis of the CODE analysis center of the IGS. High rate (30 s) satellite clock corrections were produced. A polynomial interpolation scheme was used in LEOKIN to process the 10 s observations of the spaceborne receiver for the entire day.

The maximum zenith distance of observations, which may be selected by the program user, was set to $z_{\mathrm{max}} = 90°$ for the three program runs in Table 8.16.

As already mentioned, spaceborne receivers are capable of gathering observations below the antenna's horizon. (CHAMP observations were originally made down to $z_{\max} \approx 110°$; in spring 2002 the cut-off angle was eventually set to $z = 90°$). When $z_{\max} \leq 90°$, program LEOKIN automatically weights the observations with $\cos^2 z$ to account for effects like multipath (which increase with increasing zenith distance). The absence of a priori orbit information poses particular problems for the preprocessing of observations. It is a non-trivial problem to recognize and eliminate bad code data in the absence of an a priori orbit. One may check the consistency of the pseudoranges in L_1 and L_2 and one may estimate the mean error of the observations in the adjustment (point positioning) and compare it with the a priori known value for the receiver. This latter check is, however, not very reliable, because the degree of freedom of the adjustment is only $f = n_{\text{sat}} - n_{\text{par}}$, where n_{sat} is the number of simultaneously observed GPS satellites and where $n_{\text{par}} = 4$ is the number of parameters (three coordinates and one clock term). As the blackjack receiver only tracked (at maximum) eight GPS satellites simultaneously in the time period of our example, the degree of freedom (number of observations minus number of parameters) was $f = 4$ at best – not a luxurious over-determination. 8481 satellite position vectors $\boldsymbol{r}(t_l)$ out of about 8640 possible vectors could be determined by program LEOKIN. The comparison of the resulting C1-Orbit with the reference orbit R-Orbit is not really exciting: the rms error of the Helmert transformation is about 4 m per satellite coordinate (pair).

C2-Orbit: The positions of the C1-Orbit may now be used as artificial observations by program SATORB to estimate an orbit with 15 dynamical parameters (six initial osculating elements and nine deterministic parameters) and 42 pseudo-stochastic pulses. Three stochastic pulses in the $\mathcal{R}$-system (defined in Table 4.3) were set up every 90 minutes (i.e., once per revolution). An a priori $\sigma(\delta v)$ of 2 cm was used for each pulse. The resulting C2-Orbit, established in seven orbit improvement steps, is a first simple *reduced dynamics* orbit. The characteristics of this solution are referred to as C2-Orbit (second line of Table 8.16).

As the entire day of data is treated as one arc, the a priori model for the force field has to be rather accurate in SATORB. The gravity field used to generate the results in this section is the JGM3 (see [120]), where the terms up to degree and order $n = m = 70$ were included, gravitational attractions from Sun and Moon were taken into account, the MSIS-drag model (compare section II- 3.6.1) was used, solid Earth- and ocean tides were included, as well. The relativistic version of the equations of motion were solved. Even with this considerable investment in physical and mathematical modelling one had to solve for nine empirical acceleration parameters (constant and periodic ("once per revolution") parameters) in the $\mathcal{R}$-system, plus one set of three pseudo-stochastic pulses every 90 minutes in order to obtain a fit of the positions with appropriate accuracy (i.e., with an accuracy of about one meter per coordinate).

Program SATORB allows it to screen the observations using a 3σ-criterion. This option was used for this runs. 8024 vectors (out of the 8481 vectors in the C1-Orbit) were actually used, which indicates that about 5% of the data were rejected by the screening procedure in program SATORB. Table 8.16 reveals that the reduced dynamics orbit generated by program SATORB is of much better accuracy than the input orbit C1-Orbit: instead of an rms of about 4 meters, we now have an orbit of an accuracy of about 60cm! This example underlines the power of the dynamical and reduced dynamics orbit modeling techniques. The gain of course could be achieved thanks to the greatly reduced number of unknowns (only 60 instead of $4 \cdot 8481$ in LEOKIN).

The residuals of the artificial observations (i.e., the position vectors of the satellites) in the $\mathcal{R}$-system may be inspected in Figure 8.29 (left). Observe that the residuals in radial direction are considerably (by a factor of about three) larger than the residuals in S- and W-directions. This is a well-known error characteristic for all GPS-derived positions, be they acquired on Earth or in space. Observe, that there are two short gaps of data, around $t = 1000$ min and $t = 1300$ min (time relative to the start of the day 140 in 2001 in minutes). Obviously, the screening procedure in LEOKIN experienced at times problems to screen the code observations in an appropriate way without a priori information. These problems were partly removed by program SATORB. It is amazing that the reduced dynamics orbit C2-Orbit based "only" on code-derived positions results in an orbit agreeing with the R-Orbit already on the level of about 60 cm rms per coordinate (see Table 8.16).

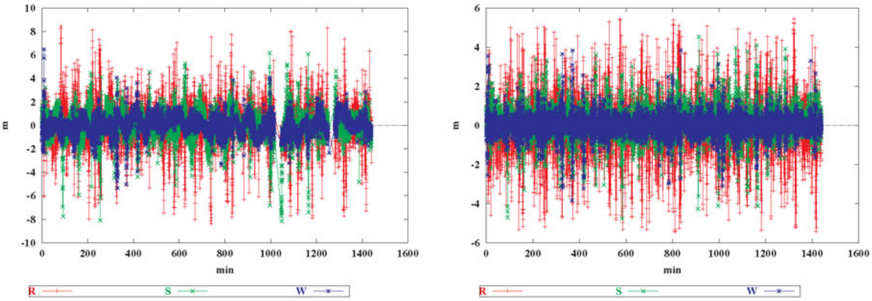


Fig. 8.29. RMS of code observations, (left: C2-Orbit, right: P2-Orbit)

C3-Orbit: The orbit resulting from the analysis using only code observations may now be used in LEOKIN to produce a better solution (screening of code observations is much easier with an orbit accurate to ≤ 1 m rms being available): Code errors of the order of more than about 5 m are easily recognized and eliminated, because the clock error is the only remaining "truly unknown" quantity in the observation eqns. (8.87). The C3-Orbit is a kinematic orbit. It only differs from the C1-Orbit by the use of a known a priori orbit. Data screening obviously was much easier in LEOKIN: The resulting rms error per coordinate (compared to the R-Orbit) dropped to the value of 1.34 m (from about four meters for the C1-Orbit). 8663 position vectors were determined by point positioning (which indicates that the data span was slightly longer than one day).

C4-Orbit: In analogy to the C2-Orbit we now produce the C4-Orbit using the position vectors in the C3-Orbit as artificial observations. As we know that the observations already are relatively clean, we may improve the force field by setting up a set of three stochastic pulses every 10 min, and we do not allow for a further cleaning of observations. The resulting reduced dynamics orbit, named C4-Orbit, is of rather good quality. An rms-error of 48 cm per coordinate results in the Helmert transformation of orbits C4 and R. All 8663 position vectors were used in the adjustment.

It is important to know that the reduced dynamics orbits resulting from SATORB, which are based only on the code observations yield orbits (C2- and C4-orbits) which already coincide to within about half a meter rms per coordinate with the best achievable LEO orbits!

P1-Orbit: With an orbit of type "C1" or "C2" at hand, position differences $\delta \boldsymbol{r}_{l,l+1} = \boldsymbol{r}_{l+1} - \boldsymbol{r}_l$ may be predicted to about one centimeter – which makes phase preprocessing an easy game, as well. With cleaned phase and code observations it is now possible to generate not only a table of positions, but also one of position differences in program LEOKIN. LEOKIN also generates a purely kinematic orbit by combining (in an extremely efficient way) the estimated positions and position differences into a table of positions. The rms error of the Helmert transformation of this purely kinematic ephemeris with the R-Orbit is only 14 cm (see Table 8.16)! It only has the disadvantage of a few data gaps (due to lack of data or due to data problems removed by screening).

P2-Orbit: The positions and position differences of the P1-Orbit are now used by program SATORB to generate a reduced dynamics orbit. The same orbit model as in the case of the C4-Orbit was used. It is interesting to compare the code residuals w.r.t. the first (C4-Orbit) and the second (P2-Orbit) orbit in Figure 8.29 (left and right). First, we observe that the overall rms of the observations is somewhat reduced (observe the scale differences). Secondly, the data gaps around 1000, 1300 min disappeared. The improvements are a consequence of the improved screening of code in program LEOKIN.

Figure 8.30 shows the residuals of the phase difference observations. Obviously, the observations are represented with an accuracy of a few centimeters only. This means in summary that the P2-Orbit is capable of representing the satellite positions to within about $1 - 2$ meters, the position differences

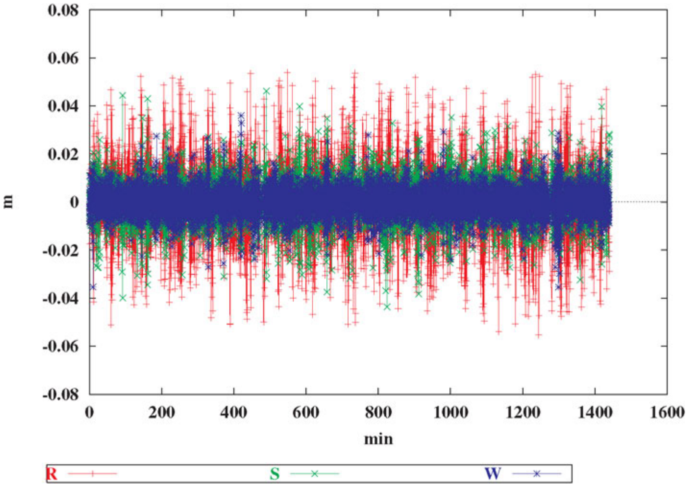


Fig. 8.30. Residuals of position differences for "P2-Orbit"

to within few centimeters. The comparison of the P2-Orbit with the R-Orbit is roughly comparable as the corresponding comparison of the P1-Orbit.

Unnecessary to say that orbits of the P1- and P2-type are highly interesting for many applications. It is interesting to note that orbits of this quality may be rather easily established!

It is of course possible to use the P2-Orbit again as a priori orbit for program LEOKIN and to repeat the entire orbit determination cycle. The achieved improvement (if observable at all) would be marginal. Much better results (one to few cm accuracy) are only achievable by a correct data processing directly relating the observations with the unknown parameters, where in particular the mathematical correlations of the phase observations in time are fully accounted for. Much more information concerning this approximate, but efficient way of establishing precise LEO orbits may be found in [24].

References

1. M. Abramowitz, I. A. Stegun: *Handbook of Mathematical Functions* (Dover Publ., New York 1965)
2. T. J. Ahrens (ed.): *Global Earth Physics – A Handbook of Physical Constants* (American Geophysical Union, Washington D.C. 1995)
3. R. R. Allen, G. E. Cook: 'The long-period motion of the plane of a distant circular orbit', Proc. R. Soc. Lond., Ser. A, **280**, 97–109, (1964)
4. E. F. Arias, P. Charlot, M. Feissel, J.-F. Lestrade: 'The extragalactic reference system of the International Earth Rotation Service, ICRS', Astron. Astrophys., **303**, 604–608 (1995)
5. R. T. H. Barnes, R. Hide, A. A. White, C. A. Wilson: 'Atmospheric angular momentum fluctuations, length-of-day changes and polar motion', Proc. R. Soc. Lond., Ser. A, **387**, 31–73 (1983)
6. A. L. Berger: 'Obliquity and Precession for the Last 5 000 000 Years', Astron. Astrophys., **51**, 127–135 (1976)
7. A. L. Berger: 'Long-Term Variations of the Earth's Orbital Elements', Cel. Mech., **15**, 53–74 (1977)
8. A. L. Berger: 'The Milankovitch Astronomical Theory of Paleoclimates: A Modern Review', Vistas in Astron., **24**, 103–122 (1980)
9. G. Beutler: *Integrale Auswertung von Satellitenbeobachtungen*, (Schweizerische Geodätische Kommission, Zürich 1977), Astronomisch-geodätische Arbeiten in der Schweiz, **33**
10. G. Beutler: *Lösung von Parameterbestimmungsproblemen in Himmelsmechanik und Satellitengeodäsie mit modernen Hilfsmitteln*, (Schweizerische Geodätische Kommission, Zürich 1982), Astronomisch-geodätische Arbeiten in der Schweiz, **34**
11. G. Beutler: *Himmelsmechanik II: Der erdnahe Raum. Mit einem Anhang von Andreas Verdun*, (Astronomisches Institut, Universität Bern, Bern 1992), Mitteilungen der Satelliten-Beobachtungsstation Zimmerwald, **28**
12. G. Beutler, E. Brockmann, W. Gurtner, U. Hugentobler, L. Mervart, M. Rothacher, A. Verdun: 'Extended orbit modeling techniques at the CODE processing center of the international GPS service for geodynamics (IGS): theory and initial results', Manuscr. Geod., **19**, 367–386 (1994)
13. G. Beutler, J. Kouba, T. Springer: 'Combining the orbits of the IGS Analysis Centers', Bull. Géod., **69**, 200–222 (1995)
14. G. Beutler: Rotation der Erde: Theorie, Methoden, Resultate aus Satellitengeodäsie und Astrometrie. Lecture Notes, Astronomical Institute, University of Bern, Bern (1997)
15. G. Beutler: Numerische Integration gewöhnlicher Differentialgleichungssysteme. Lecture Notes, Astronomical Institute, University of Bern, Bern (1998)
16. G. Beutler: Himmelsmechanik des Planetensystems. Lecture Notes, Astronomical Institute, University of Bern, Bern (1999)

17. G. Beutler, M. Rothacher, S. Schaer, T. A. Springer, J. Kouba, R. E. Neilan: 'The International GPS Service (IGS): An interdisciplinary service in support of Earth sciences', Adv. Space Res., **23**, 631–653 (1999)
18. G. Beutler: Himmelsmechanik des erdnahen Raumes. Lecture Notes, Astronomical Institute, University of Bern, Bern (2000)
19. G. Beutler, M. Rothacher, J. Kouba, R. Weber: 'Polar Motion with Daily and Sub-daily Time Resolution'. In: *Polar Motion: Historical and Scientific Problems, IAU Colloquium 178, Cagliari, Sardinia, Italy, 27–30 September 1999*, ed. by S. Dick, D. McCarthy, B. Luzum (Astronomical Society of the Pacific, San Francisco 2000), ASP Conference Series, **208**, pp. 513–525
20. J. Binney, S. Tremaine: *Galactic Dynamics* (Princeton University Press, Princeton 1987)
21. H. Bock, G. Beutler, S. Schaer, T. A. Springer, M. Rothacher: 'Processing aspects related to permanent GPS arrays', Earth Planets Space, **52**, 657–662 (2000)
22. H. Bock, U. Hugentobler, T. S. Springer, G. Beutler: 'Efficient Precise Orbit Determination of LEO Satellites Using GPS', Adv. Space Res., **30**, 295–300 (2002)
23. H. Bock, G. Beutler, U. Hugentobler: 'Kinematic Orbit Determination for Low Earth Orbiter (LEOs)'. In: *Vistas for Geodesy in the New Millennium – IAG 2001 Scientific Assembly, Budapest, Hungary, September 2–7, 2001*, ed. by J. Ádám, K.-P. Schwarz (Springer, Berlin, Heidelberg 2002), International Association of Geodesy Symposia, **125**, pp. 303–308
24. H. Bock: Efficient Methods for Determining Precise Orbits of Low Earth Orbiters Using the Global Positioning System, Ph.D. Thesis, Astronomical Institute, University of Bern, Bern (2003)
25. I. N. Bronstein, K. A. Semendjajew, G. Musiol, H. Mühlig: *Taschenbuch der Mathematik*, 5th edn. (Harri Deutsch, Thun, Frankfurt/Main 2000)
26. D. Brouwer: 'On the accumulation of errors in numerical integration', Astron. J., **46**, 149–153 (1937)
27. D. Brouwer, G. M. Clemence: *Methods of Celestial Mechanics*, 2nd impr. (Academic Press, Orlando, San Diego, New York 1985)
28. H. E. Coffey (ed.): 'Data for November and December 1999', Solar-Geophys. Data, **665**, Part I, 127–135 (2000)
29. C. J. Cohen, E. C. Hubbard, C. Oesterwinter: *Elements Of The Outer Planets For One Million Years*, (The Nautical Almanac Office, U.S. Naval Observatory, Washington D.C. 1973) Astronomical Papers Prepared for the Use of the American Ephemeris and Nautical Almanac, **22**, Part 1
30. C. J. Cohen, E. C. Hubbart, C. Oesterwinter: 'Planetary elements for 10 000 000 years', Cel. Mech., **7**, 438–448 (1973)
31. J. M. A. Danby: *Fundamentals of Celestial Mechanics*, 2nd edn. (Willmann-Bell, Richmond, Virginia 1989)
32. A. T. Doodson: 'The harmonic development of the tide-generating potential', Proc. R. Soc. Lond., Ser. A, **100**, 305–329 (1922)
33. L. Euler: 'Recherches sur le mouvement des corps célestes en général', Hist. l'Acad. Roy. Sci. et belles lettres (Berlin), **3**, 93–143 (1749); E. 112, O. II, **25**, 1–44 (1960)
34. L. Euler: 'Découverte d'un nouveau principe de mécanique', Hist. l'Acad. Roy. Sci. et belles lettres (Berlin), **6**, 185–217 (1752); E. 177, O. II, **5**, 81–108 (1957)
35. L. Euler: 'Principes généraux du mouvement des fluides', Hist. l'Acad. Roy. Sci. et belles lettres (Berlin), **11**, 274–315 (1757); E. 226, O. II, **12**, 54–91 (1954)

36. L. Euler: ‘Du mouvement de rotation des corps solides autour d’un axe variable’. Hist. l’Acad. Roy. Sci. et belles lettres (Berlin), **14**, 154–193 (1765); E. 292, O. II, **8**, 200–235 (1965)
37. L. Euler: *Institutionum calculi integralis volumen primum in quo methodus integrandi a primis principiis usque ad integrationem aequationum differentialium primi gradus pertractatur* (Acad. imp. sc., Petropoli 1768), pp. 493–508; E. 342, O. I, **11**, 424–434 (1913)
38. E. Fehlberg: Classical fifth-,sixth-, seventh-, and eight-order Runge-Kutta formulas with stepsize control. NASA Technical Report R-287 (NASA, Huntsville 1968)
39. S. Ferras-Mello, T. A. Michtchenko, D. Nesvorný, F. Roig, A. Simula: ‘The depletion of the Hecuba gap vs the long-lasting Hilda group’, Planet. Space Sci., **46**, 1425–1432 (1998)
40. H. F. Fliegel, T. E. Galini, E. R. Swift: ‘Global Positioning System Radiation Force Model for Geodetic Applications’, J. Geophys. Res., **97**, 559–568 (1992)
41. W. Flury: Raumfahrtmechanik. Lecture Notes, Technical University of Darmstadt, Darmstadt (1994)
42. C. Froeschlé, Ch. Froeschlé: ‘Order and chaos in the solar system’. In: *Proceedings of the Third International Workshop on Positional Astronomy and Celestial Mechanics, University of Valencia, Cuenca, Spain, October 17–21, 1994*, ed. by A. López Garcia, E. I. Yagudina, M. J. Martinez Usó, A. Cordero Barbero (Universitat de Valencia, Observatorio Astronomico, Valencia 1996) pp. 155–171
43. L.-L. Fu, E. J. Christensen, C. A. Yamarone, M. Lefebvre, Y. Ménard, M. Dorrer, P. Escudier: ‘TOPEX/POSEIDON mission overview’, J. Geophys. Res. **99**, 24369–24381 (1994)
44. C. F. Gauss: ‘Summarische Übersicht der zur Bestimmung der Bahnen der beyden neuen Hauptplaneten angewandten Methoden’, Monatl. Corresp., **20**, 197–224 (1809)
45. C. W. Gear: *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs, New Jersey 1971)
46. J. M. Gipson: ‘Very long baseline interferometry determination of neglected tidal terms in high-frequency Earth orientation variation’, J. Geophys. Res., **101**, 28051–28064 (1996)
47. W. B. Gragg: ‘On extrapolation algorithms for ordinary initial value problems’, J. SIAM, Ser. B, **2**, 384–403 (1965)
48. W. Gurtner: ‘RINEX: The Receiver-Independent Exchange Format’, GPS World, **5**, No. 7, 48–52 (1994)
49. A. Guthmann: *Einführung in die Himmelsmechanik und Ephemeridenrechnung – Theorie, Algorithmen, Numerik*, 2. Aufl. (Spektrum Akademischer Verlag, Heidelberg, Berlin, 2000)
50. A. E. Hedin: ‘MSIS-86 Model’, J. Geophys. Res., **92**, 4649–4662 (1987)
51. A. E. Hedin: ‘Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere’, J. Geophys. Res., **96**, 1159–1172 (1991)
52. W. A. Heiskanen, H. Moritz: *Physical Geodesy* (W. H. Freeman Comp., San Francisco, London 1967)
53. P. Henrici: *Discrete Variable Methods in Ordinary Differential Equations* (John Wiley & Sons, New York, London Sidney 1968)
54. P. Herget: ‘Computation of Preliminary Orbits’, Astron. J., **70**, 1–3 (1965)
55. T. A. Herring: ‘An a priori model for the reduction of nutation observations: $\mathrm{KSV}_{1994.3}$ nutation series’, Highlights of Astronomy, **10**, 222–227 (1995)
56. K. Hirayama: ‘Groups of Asteroids Probably of Common Origin’, Astron. J., **31**, 185–188 (1918)

57. U. Hugentobler: *Astrometry and Satellite Orbits: Theoretical Considerations and Typical Applications*, (Schweizerische Geodätische Kommission, Zürich 1998), Geodätisch-geophysikalische Arbeiten in der Schweiz, **57**
58. U. Hugentobler, S. Schaer, P. Fridez: *Bernese GPS Software Version 4.2* (Astronomical Institute, University of Bern, Bern 2001)
59. U. Hugentobler, D. Ineichen, G. Beutler: 'GPS satellites: Radiation pressure, attitude and resonance', Adv. Space Res. **31**, 1917–1926 (2003)
60. J. Imbrie: 'Astronomical Theory of the Pleistocene Ice Ages: A Brief Historical Review', Icarus, **50**, 408–422 (1982)
61. A. Jäggi: 'Efficient Stochastic Orbit Modeling Techniques using Least Squares Estimators', International Association of Geodesy Symposia, **127** (in press)
62. W. M. Kaula: *Theory of Satellite Geodesy – Applications of Satellites to Geodesy* (Blaisdell Publ. Comp., Waltham, Toronto, London 1966)
63. Z. Kopal: *Numerical Analysis* (Chapman & Hall, London 1955)
64. J. Kouba: 'A review of geodetic and geodynamic satellite Doppler positioning', Rev. Geophys. Space Phys., **21**, 27–40 (1983)
65. J. Kouba, G. Beutler, M. Rothacher: 'IGS Combined and Contributed Earth Rotation Parameter Solutions'. In: *Polar Motion: Historical and Scientific Problems, IAU Colloquium 178, Cagliari, Sardinia, Italy, 27–30 September 1999*, ed. by S. Dick, D. McCarthy, B. Luzum (Astronomical Society of the Pacific, San Francisco 2000), ASP Conference Series, **208**, pp. 277–302
66. K. Lambeck: *Geophysical Geodesy – The Slow Deformations of the Earth* (Clarendon Press, Oxford 1988)
67. L D. Landau, J. M. Lifschitz: *Lehrbuch der theoretischen Physik*, Bd. 6 (Hydrodynamik), 5. Aufl., (Verlag Harri Deutsch, Frankfurt/Main 1991)
68. K. R. Lang: *Astrophysical Data: Planets and Stars* (Springer, New York, Berlin, Heidelberg 1992)
69. W. Lowrie: *Fundamentals of Geophysics* (Cambridge University Press, Cambridge 1997)
70. D. D. McCarthy: *IERS Conventions (1996)*, (Central Bureau of IERS, Observatoire de Paris, Paris 1996) IERS Technical Note, **21**
71. D. D. McCarthy: *IERS Conventions (2000)*, (Central Bureau of IERS, Observatoire de Paris, Paris 2004) IERS Technical Note, **32** (in press)
72. J. Meeus: *Astronomical Algorithms*, 2nd edn. (Willmann-Bell, Richmond, Virginia 1999)
73. W. G. Melbourne, E. S. Davis, C. B. Duncan, G. A. Hajj, K. R. Hardy, E. R. Kursinski, T. K. Meehan, L. E. Young, T. P. Yunck: *The Application of Spaceborne GPS to Atmospheric Limb Sounding and Global Change Monitoring* (NASA, JPL, Pasadena, 1994), JPL Publication 94-18
74. A. Milani, A. M. Nobili, P. Farinella: *Non-Gravitational Perturbations and Satellite Geodesy* (Adam Hilger, Bristol 1987)
75. O. Montenbruck, E. Gill: *Satellite Orbits – Models, Methods, and Appications*, corr. 2nd. print. (Springer, Berlin, Heidelberg 2001)
76. H. Moritz, I. I. Mueller: *Earth Rotation – Theory and Observation* (Ungar Publ. Comp., New York 1988)
77. F. R. Moulton: *An Introduction to Celestial Mechanics*, 2nd rev. edn. (Dover Publ., New York 1970)
78. W. H. Munk, G. J. F. Macdonald: *The Rotation of the Earth – A Geophysical Discussion*, 2nd edn. (Cambridge University Press, Cambridge 1975)
79. P. Murdin (ed.): *Encyclopedia of Astronomy and Astrophysics* (Institute of Physics Publ., Bristol, Philadelphia 2001)
80. C. D. Murray, S. F. Dermott: *Solar System Dynamics* (Cambridge University Press, Cambridge 1999)

81. NAg-Library: *The NAG Fortran Library Introductory Guide, Mark 17* (NAG-Ltd, Wilkinson House, Oxford 1995)
82. X. X. Newhall, E. M. Standish, J. G. Williams: 'DE 102: a numerically integrated ephemeris of the Moon and planets spanning forty-four centuries', Astron. Astrophys., **125**, 150–167 (1983)
83. I. Newton: *Philosophiae naturalis principia mathematica* (Jussu Societatis Regiae ac Typis Josephi Streater, Londini 1687)
84. I. Newton: *The Principia – Mathematical Principles of Natural Philosophy*, Transl. by I. B. Cohen and A. Whitman (University of California Press, Berkeley, Los Angeles, London 1999)
85. I. Peterson: *Newton's Clock – Chaos in the Solar System* (W. H. Freeman Comp., New York 1993)
86. H. Poincaré: *New Methods of Celestial Mechanics*, ed. by D. L. Goroff (American Institute of Physics, USA, 1993)
87. H. Poincaré: 'Sur la précession des corps déformables', Bull. Astron. (Paris), **27**, 321–356 (1910)
88. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: *Numerical Recipes in Fortran 77 – The Art of Scientific Computing*, 2nd edn. (Cambridge University Press, Cambridge 1996)
89. C. Reigber: 'Gravity Field Recovery from Satellite Tracking Data'. In: *Theory of Satellite Geodesy and Gravity Field Determination*, ed. by F. Sansò, R. Rummel (Springer, Berlin, Heidelberg 1989), Lecture Notes in Earth Sciences, **25**, pp. 197–234
90. F. P J. Rimrott: *Introductory Orbit Dynamics* (Vieweg & Sohn, Braunschweig, Wiesbaden 1989)
91. N. T. Roseveare: *Mercury's perihelion from Le Verrier to Einstein* (Clarendon Press, Oxford 1982)
92. M. Rothacher, G. Beutler, T. A. Herring, R. Weber: 'Estimation of nutation using the Global Positioning System' J. Geophys. Res., **104**, 4835–4859 (1999)
93. M. Rothacher, G. Beutler, R. Weber, J. Hefty: 'High-frequency variations in Earth rotation from Global Positioning System data', J. Geophys. Res., **106**, 13711–13738 (2001)
94. A. E. Roy: *Orbital Motion*, 3rd edn. (Adam Hilger, Bristol, Philadelphia 1988)
95. A. E. Roy, I. W. Walker, A. J. Macdonald, I. P. Williams, K. Fox, C. D. Murray, A. Milani, A. M. Nobili, P. J. Message, A. T. Sinclair, M. Carpino: 'Project LONGSTOP', Vistas in Astron., **32**, 95–116 (1988)
96. D. A. Salstein, D. M. Kann, A. J. Miller, R. D. Rosen: 'The Sub-bureau for Atmospheric Angular Momentum of the International Earth Rotation Service: A Meteorological Data Center with Geodetic Applications', Bull. Amer. Meteor. Soc., **74**, 67–80 (1993)
97. S. Schaer, G. Beutler, M. Rothacher: 'Mapping and Predicting the Ionosphere'. In: *Proceedings of the IGS 1998 Analysis Center Workshop, ESA/ESOC, Darmstadt, Germany, February 9–11, 1998*, ed. by. J. M. Dow, J. Kouba, T. Springer (ESA/ESOC, Darmstadt 1998) pp. 307–318
98. S. Schaer: *Mapping and Predicting the Earth's Ionosphere Using the Global Positioning System*, (Schweizerische Geodätische Kommission, Zürich 1999), Geodätisch-geophysikalische Arbeiten in der Schweiz, **59**
99. T. Schildknecht, U. Hugentobler, M. Ploner: 'Optical surveys of space debris in GEO', Adv. Space Res., **23**, 45–54 (1999)
100. T. Schildknecht, M. Ploner, U. Hugentobler: 'The search for debris in GEO', Adv. Space Res., **28**, 1291–1299 (2001)
101. T. Schildknecht, R. Musci, M. Ploner, S. Preisig, J. de Leon Cruz, H. Krag: 'Optical Observation of Space Debris in the Geostationary Ring'. In: *Proceed-*

ings of the Third European Conference on Space Debris, March 19-21, 2001, ESOC, Darmstadt, Germany, (ESA Publ. Div., ESTEC, Noordwijk 2001), SP-473, Vol. 1, pp. 89–93

102. L. D. Schmadel: *Dictionary of Minor Planet Names*, 4th rev. and enl. edn. (Springer, Berlin, Heidelberg 1999)
103. M. Schneider: *Satellitengeodäsie* (B.I. Wissenschaftsverlag, Mannheim, Wien, Zürich 1988)
104. J. Schubart: 'Three Characteristic Parameters of Orbits of Hilda-type Asteroids', Astron. Astrophys., **114**, 200–204 (1982)
105. J. Schubart: 'Additional results on orbits of Hilda-type asteroids', Astron. Astrophys., **241**, 2997–302 (1991)
106. G. Seeber: *Satellite Geodesy – Foundations, Methods, and Applications*, 2nd edn. (Walter de Gruyter, Berlin, New York 2003)
107. P. K. Seidelmann (ed.): *Explanatory Supplement to the Astronomical Almanac* (University Science Books, Mill Valley, California, 1992)
108. L. F. Shampine, M. K. Gordon: *Computer Solution of Ordinary Differential Equations – The Initial Value Problem*, (W. H. Freeman Comp., San Francisco 1975)
109. M. H. Soffel: *Relativity in Astrometry, Celestial Mechanics and Geodesy* (Springer, Berlin, Heidelberg 1989)
110. T. A. Springer: *Modeling and Validating Orbits and Clocks Using the Global Positioning System*, (Schweizerische Geodätische Kommission, Zürich 2000), Geodätisch-geophysikalische Arbeiten in der Schweiz, **60**
111. E. M. Standish: 'The observational basis for JPL's DE 200, the planetary ephemerides of the Astronomical Almanac', Astron. Astrophys., **233**, 252–271 (1990)
112. J. Stoer, R. Bulirsch: *Einführung in die Numerische Mathematik*, 2. Aufl. (Springer, Berlin, Heidelberg 1976-1978), Heidelberger Taschenbücher, **105, 114**
113. J. Stoer, R. Bulirsch: *Introduction to numerical analysis*, 3rd edn., (Springer, New York 2002), Texts in applied mathematics, **12**
114. K. Stumpff: *Himmelsmechanik* (VEB Deutscher Verlag der Wissenschaften, Berlin 1959–1974)
115. D. Švehla, M. Rothacher: 'Kinematic Orbit Determination of LEOs Based on Zero or Double-difference Algorithms Using Simulated and Real SST GPS Data'. In: *Vistas for Geodesy in the New Millennium – IAG 2001 Scientific Assembly, Budapest, Hungary, September 2–7, 2001*, ed. by J. Ádám, K.-P. Schwarz (Springer, Berlin, Heidelberg 2002), International Association of Geodesy Symposia, **125**, pp. 322–328
116. D. Švehla, M. Rothacher: 'CHAMP Double-Difference Kinematic POD with Ambiguity Resolution'. In: *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, ed. by C. Reigber, H. Lühr, P. Schwintzer (Springer, Berlin, Heidelberg 2003) pp. 70–77
117. V. Szebehely: *Theory of Orbits – The Restricted Problem of Three Bodies* (Academic Press, New York, London 1967)
118. L. G. Taff: *Celestial Mechanics – A Computational Guide for the Practitioner* (John Wiley & Sons, New York, Chichester 1985)
119. B. D. Tapley: 'Fundamentals of Orbit Determination'. In: *Theory of Satellite Geodesy and Gravity Field Determination*, ed. by F. Sansò, R. Rummel (Springer, Berlin, Heidelberg 1989), Lecture Notes in Earth Sciences, **25**, pp. 235–260
120. B. D. Tapley, M. M. Watkins, J. C. Ries, G. W. Davies, R. J. Eanes, S. R. Poole, H. J. Rim, B. E. Schutz, C. K. Shum, R. S. Nerem, F. J. Lerch,

J. A. Marshall, S. M. Klosko, N. K. Pavlis, R. G. Williamson: 'The Joint Gravity Model 3', J. Geophys. Res., **101**, 28029–28049 (1996)

121. F. Tissérand: *Traité de Mécanique Céleste*, Nouv. tirage (Gauthier-Villars, Paris 1960)
122. P. J. G. Teunissen, A. Kleusberg (eds.): *GPS for Geodesy*, 2nd edn., (Springer, Berlin, Heidelberg 1998)
123. 'TOPEX/POSEIDON: Geophysical Evaluation', J. Geophys. Res., **99**, 24369–25062 (1994)
124. W. Torge: *Geodesy*, 3rd edn., (Walter de Gruyter, Berlin, New York, 2001)
125. W. Torge: *Geodäsie*, 2. Aufl., (Walter de Gruyter, Berlin, New York 2003)
126. A. Verdun, G. Beutler: 'Early Observational Evidence of Polar Motion'. In: *Polar Motion: Historical and Scientific Problems, IAU Colloquium 178, Cagliari, Sardinia, Italy, 27–30 September 1999*, ed. by S. Dick, D. McCarthy, B. Luzum (Astronomical Society of the Pacific, San Francisco 2000), ASP Conference Series, **208**, pp. 67–81
127. H. G. Walter, O. J. Sovers: *Astrometry of Fundamental Catalogues – The Evolution from Optical to Radio Reference Frames* (Springer, Berlin, Heidelberg, New York 2000)
128. R. Weber, M. Rothacher: 'The Quality of Sub-daily Polar Motion Estimates based on GPS Observations'. In: *Polar Motion: Historical and Scientific Problems, IAU Colloquium 178, Cagliari, Sardinia, Italy, 27–30 September 1999*, ed. by S. Dick, D. McCarthy, B. Luzum (Astronomical Society of the Pacific, San Francisco 2000), ASP Conference Series, **208**, pp. 527–532
129. A. Wegener: *Die Entstehung der Kontinente und Ozenane* (Friedrich Vieweg & Sohn, Braunschweig 1915), Sammlung Vieweg, **23**
130. J. Wisdom: 'Chaotic Behavior and the Origin of the 3/1 Kirkwood Gap', Icarus, **56**, 51–74 (1983)
131. J. Wisdom: 'Chaotic behavior in the solar system'. Nucl. Phys. B (Proc. Suppl.), **2**, 391–414 (1987)
132. E. W. Woolard: *Theory of the Rotation of the Earth around its Center of Mass* (The Nautical Almanac Office, U.S. Naval Observatory, Washington D.C. 1953) Astronomical Papers Prepared for the Use of the American Ephemeris and Nautical Almanac, **15**, Part 1
133. V. N. Zharkov, S. M. Molodensky, A. Brzeziński, E. Groten, P. Varga: *The Earth and its Rotation – Low Frequency Geodynamics* (H. Wichmann Verlag, Heidelberg 1996)
134. J. F. Zumberge, M. B. Heflin, D. C. Jefferson, M. M. Watkins, F. H. Webb: 'Precise point positioning for the efficient and robust analysis of GPS data from large networks', J. Geophys. Res., **102**, 5005–5017 (1997)

Abbreviations and Acronyms

AAM	Atmospheric Angular Momentum
AIUB	Astronomical Institute, University of Bern
AMF	Angular Momentum Functions
AU	Astronomical Unit
BIH	Bureau International de l'Heure
BKG	Bundesamt für Kartographie und Geodäsie
CCD	Charge Coupled Device
CHAMP	CHAllenging Minisatellite Payload
CIRA	COSPAR International Reference Atmosphere
CODE	Center for Orbit Determination in Europe
COSPAR	Committee on Space Research
CPU	Central Processing Unit
CSTG	Commission on Coordination of Space Techniques
ΔLOD	Excess LOD
DE200	Development Ephemeris 200
DORIS	Doppler Orbitography by Radiopositioning Integrated on Satellite
DoY	Day of Year
ECMWF	European Center for Medium-Range Weather Forecasts
EOP	Earth Orientation Parameter
ERP	Earth Rotation Parameters
ERS-2	Earth Remote Sensing 2
ESA	European Space Agency
ET	Ephemeris Time
FCN	Free Core Nutation
FFT	Fast Fourier Transformation
FT	Fourier Transformation
GARP	Global Atmospheric Research Program
GFZ	GeoForschungsZentrum
GOCE	Gravity field and steady-state Ocean Circulation Experiment
GPS	Global Positioning System
GPS/MET	GPS Meteorology using limb sounding

GRACE	Gravity Recover and Climate Experiment
HIPPARCOS	HIgh Precision PARallax COllecting Satellite
IAG	International Association of Geodesy
IAU	International Astronomical Union
ICRF	International Celestial Reference Frame
ICRS	International Celestial Reference System
IERS	International Earth Rotation and Reference Systems Service
IGN	Institut Géographique National
IGS	International GPS Service
ILRS	International Laser Ranging Service
ILS	International Latitude Service
IPMS	International Polar Motion Service
ITRF	International Terrestrial Reference Frame
ITRS	International Terrestrial Reference System
IVS	International VLBI Service for Astrometry and Geodesy
JD	Julian Date
JGM3	Joint Gravity Model 3
JPL	Jet Propulsion Laboratory
LAGEOS	LAser GEOdetic Satellite
LEO	Low Earth Orbiter
LLR	Lunar Laser ranging
LOD	Length of Day
LSQ	Least Squares
Laser	Light Amplification through Stimulated Emission of Radiation
mas	milliarcseconds
μs/day	microseconds per day
MJD	Modified Julian Date
MPC	Minor Planet Center
MSIS	Mass Spectrometer and Incoherent Scatter
NCEP	U.S. National Centers for Environmental Prediction
NDFW	Nearly-Diurnal Free Wobble
NEA	Near Earth Asteroids
NEO	Near Earth Objects
NNSS	U.S. Navy Navigation Satellite System
NOAA	National Oceanic and Atmospheric Administration
ns	nanoseconds
PAGEOS	PAssive GEOdetic Satellite
PAI	Principal Axes of Inertia
PC	Personal Computer
PM	Polar Motion
POD	Precise Orbit Determination

PPN	Parametrized Post-Newtonian
PPP	Precise Point Positioning
PRARE	Precise Range And Range-rate Equipment
Quasars	Quasistellar Radio Sources
RINEX	Receiver Independent Exchange Format
SI	International System of units
SLR	Satellite Laser Ranging
ST	Sidereal Time
swisstopo	Swiss Federal office of Topography
TAI	International Atomic Time
TDB	Barycentric Dynamical Time
TNO	Trans-Neptunian Objects
TOPEX	TOPography EXperiment for Ocean Circulation
TT	Terrestrial Time
TUM	Technical University of Munich
UT	Universal Time
UT1	UT corrected for polar motion effects
UTC	Universal Time Coordinated
VLBI	Very Long Baseline Interferometry
WGS-84	World Geodetic System 1984

Name Index

Subject Index

Printing: Saladruck, Berlin
Binding: Stein+Lehmann, Berlin